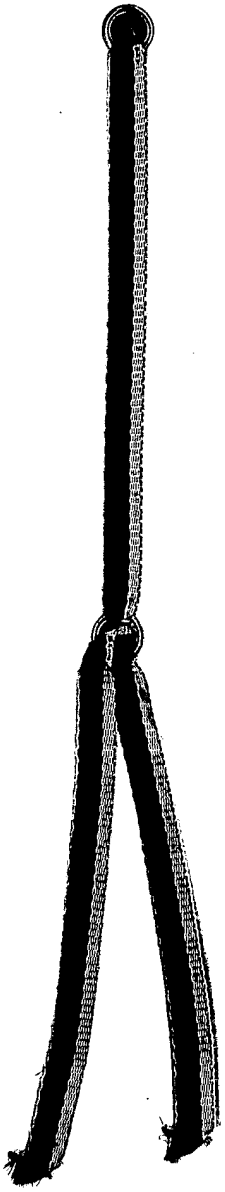
Prioritätsbescheinigung über die Einreichung einer Patentanmeldung



Aktenzeichen: 199 32 230.9

Anmeldetag: 9. Juli 1999

Anmelder/Inhaber: BASF Aktiengesellschaft, Ludwigshafen/DE

Bezeichnung: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING PROTEINS INVOLVED IN MEMBRANE SYNTHESIS AND MEMBRANE TRANSPORT

IPC: C 07 K 14/34

Die angehefteten Stücke sind eine richtige und genaue Wiedergabe der ursprünglichen Unterlagen dieser Patentanmeldung.

München, den 24. Juli 2000
Deutsches Patent- und Markenamt
Der Präsident
Im Auftrag

Weihmayr

A 9161
06/00
EDV-L

CERTIFIED COPY OF
PRIORITY DOCUMENT

BUNDESREPUBLIK DEUTSCHLAND

Prioritätsbescheinigung über die Einreichung einer Patentanmeldung

Aktenzeichen:	199 32 230.9
Anmeldetag:	9. Juli 1999
Anmelder/Inhaber:	BASF Aktiengesellschaft, Ludwigshafen/DE
Bezeichnung:	CORYNEBACTERIUM GLUTAMICUM GENES ENCODING PROTEINS INVOLVED IN MEMBRANE SYNTHESIS AND MEMBRANE TRANSPORT
IPC:	C 07 K 14/34

Die angehefteten Stücke sind eine richtige und genaue Wiedergabe der ursprünglichen Unterlagen dieser Patentanmeldung.

München, den 24. Juli 2000
Deutsches Patent- und Markenamt
Der Präsident
Im Auftrag

Weihmayr

A 9161
06/00
EDV-L

CORYNEBACTERIUM GLUTAMICUM GENES ENCODING PROTEINS INVOLVED IN MEMBRANE SYNTHESIS AND MEMBRANE TRANSPORT

Background of the Invention

Certain products and by-products of naturally-occurring metabolic processes in cells have utility in a wide array of industries, including the food, feed, cosmetics, and pharmaceutical industries. These molecules, collectively termed 'fine chemicals', include organic acids, both proteinogenic and non-proteinogenic amino acids, nucleotides and nucleosides, lipids and fatty acids, diols, carbohydrates, aromatic compounds, vitamins and cofactors, and enzymes. Their production is most conveniently performed through the large-scale culture of bacteria developed to produce and secrete large quantities of one or more desired molecules. One particularly useful organism for this purpose is *Corynebacterium glutamicum*, a gram positive, nonpathogenic bacterium. Through strain selection, a number of mutant strains have been developed which produce an array of desirable compounds. However, selection of strains improved for the production of a particular molecule is a time-consuming and difficult process.

Summary of the Invention

This invention provides novel nucleic acid molecules which may be used to identify or classify *Corynebacterium glutamicum* or related species of bacteria. *C. glutamicum* is a gram positive, aerobic bacterium which is commonly used in industry for the large-scale production of a variety of fine chemicals, and also for the degradation of hydrocarbons (such as in petroleum spills) and for the oxidation of terpenoids. The nucleic acid molecules therefore can be used to identify microorganisms which can be used to produce fine chemicals, e.g., by fermentation processes. While *C. glutamicum* itself is nonpathogenic, it is related to other Corynebacterium species, such as *Corynebacterium diphtheriae* (the causative agent of diphtheria), which are important human pathogens. The ability to identify the presence of Corynebacterium species therefore also can have significant clinical relevance, e.g., diagnostic applications. Further, these nucleic acid molecules may serve as reference points for the mapping of the *C. glutamicum* genome, or of genomes of related organisms.

These novel nucleic acid molecules encode proteins, referred to herein as membrane construction and membrane transport (MCT) proteins. These MCT proteins are capable of, for example, performing a function involved in the metabolism (e.g., the biosynthesis or degradation) of compounds necessary for membrane biosynthesis, or of assisting in the transmembrane transport of one or more compounds either into or out of

the cell. Given the availability of cloning vectors for use in *Corynebacterium glutamicum*, such as those disclosed in Sinskey et al., U.S. Patent No. 4,649,119, and techniques for genetic manipulation of *C glutamicum* and the related *Brevibacterium* species (e.g., *lactofermentum*) (Yoshihama et al, *J. Bacteriol.* 162: 591-597 (1985); Katsumata et al., *J. Bacteriol.* 159: 306-311 (1984), and Santamaria et al., *J. Gen. Microbiol.* 130: 2237-2246 (1984)), the nucleic acid molecules of the invention may be utilized in the genetic engineering of this organism to make it a better or more efficient producer of one or more fine chemicals. This improved production or efficiency of production of a fine chemical may be due to a direct effect of manipulation of a gene of the invention, or it may be due to an indirect effect of such manipulation.

There are a number of mechanisms by which the alteration of an MCT protein of the invention may directly affect the yield, production, and/or efficiency of production of a fine chemical from a *C. glutamicum* strain incorporating such an altered protein. Those MCT proteins involved in the export of fine chemical molecules from the cell may be increased in number or activity such that greater quantities of these compounds are secreted to the extracellular medium, from which they are more readily recovered. Similarly, those MCT proteins involved in the import of nutrients necessary for the biosynthesis of one or more fine chemicals (e.g., phosphate, sulfate, nitrogen compounds, etc.) may be increased in number or activity such that these precursors, cofactors, or intermediate compounds are increased in concentration within the cell. Further, fatty acids and lipids themselves are desirable fine chemicals; by optimizing the activity or increasing the number of one or more MCT proteins of the invention which participate in the biosynthesis of these compounds, or by impairing the activity of one or more MCT proteins which are involved in the degradation of these compounds, it may be possible to increase the yield, production, and/or efficiency of production of fatty acid and lipid molecules from *C. glutamicum*.

The mutagenesis of one or more MCT genes of the invention may also result in MCT proteins having altered activities which indirectly impact the production of one or more desired fine chemicals from *C .glutamicum*. For example, MCT proteins of the invention involved in the export of waste products may be increased in number or activity such that the normal metabolic wastes of the cell (possibly increased in quantity due to the overproduction of the desired fine chemical) are efficiently exported before they are able to damage nucleotides and proteins within the cell (which would decrease the viability of the cell) or to interfere with fine chemical biosynthetic pathways (which would decrease the yield, production, or efficiency of production of the desired fine chemical). Further, the relatively large intracellular quantities of the desired fine chemical may in itself be toxic to the cell, so by increasing the activity or number of

transporters able to export this compound from the cell, one may increase the viability of the cell in culture, in turn leading to a greater number of cells in the culture producing the desired fine chemical. The MCT proteins of the invention may also be manipulated such that the relative amounts of different lipid and fatty acid molecules are produced. This may have a profound effect on the lipid composition of the membrane of the cell. Since each type of lipid has different physical properties, an alteration in the lipid composition of a membrane may significantly alter membrane fluidity. Changes in membrane fluidity can impact the transport of molecules across the membrane, as well as the integrity of the cell, both of which have a profound effect on the production of fine chemicals from *C. glutamicum* in large-scale fermentative culture.

The invention provides novel nucleic acid molecules which encode proteins, referred to herein as MCT proteins, which are capable of, for example, participating in the metabolism of compounds necessary for the construction of cellular membranes in *C. glutamicum*, or in the transport of molecules across these membranes. Nucleic acid molecules encoding an MCT protein are referred to herein as MCT nucleic acid molecules. In a preferred embodiment, the MCT protein participates in the metabolism of compounds necessary for the construction of cellular membranes in *C. glutamicum*, or in the transport of molecules across these membranes. Examples of such proteins include those encoded by the genes set forth in Table 1.

Accordingly, one aspect of the invention pertains to isolated nucleic acid molecules (e.g., cDNAs) comprising a nucleotide sequence encoding an MCT protein or biologically active portions thereof, as well as nucleic acid fragments suitable as primers or hybridization probes for the detection or amplification of MCT-encoding nucleic acid (e.g., DNA or mRNA). In particularly preferred embodiments, the isolated nucleic acid molecule comprises one of the nucleotide sequences set forth in Appendix A or the coding region or a complement thereof of one of these nucleotide sequences. In other particularly preferred embodiments, the isolated nucleic acid molecule of the invention comprises a nucleotide sequence which hybridizes to or is at least about 50%, preferably at least about 60%, more preferably at least about 70%, 80% or 90%, and even more preferably at least about 95%, 96%, 97%, 98%, 99% or more homologous to a nucleotide sequence set forth in Appendix A, or a portion thereof. In other preferred embodiments, the isolated nucleic acid molecule encodes one of the amino acid sequences set forth in Appendix B. The preferred MCT proteins of the present invention also preferably possess at least one of the MCT activities described herein.

In another embodiment, the isolated nucleic acid molecule encodes a protein or portion thereof wherein the protein or portion thereof includes an amino acid sequence which is sufficiently homologous to an amino acid sequence of Appendix B, e.g.,

sufficiently homologous to an amino acid sequence of Appendix B such that the protein or portion thereof maintains an MCT activity. Preferably, the protein or portion thereof encoded by the nucleic acid molecule maintains the ability to participate in the metabolism of compounds necessary for the construction of cellular membranes in *C. glutamicum*, or in the transport of molecules across these membranes. In one embodiment, the protein encoded by the nucleic acid molecule is at least about 50%, preferably at least about 60%, and more preferably at least about 70%, 80%, or 90% and most preferably at least about 95%, 96%, 97%, 98%, or 99% or more homologous to an amino acid sequence of Appendix B (e.g., an entire amino acid sequence selected from those sequences set forth in Appendix B). In another preferred embodiment, the protein is a full length *C. glutamicum* protein which is substantially homologous to an entire amino acid sequence of Appendix B (encoded by an open reading frame shown in Appendix A).

In another preferred embodiment, the isolated nucleic acid molecule is derived from *C glutamicum* and encodes a protein (e.g., an MCT fusion protein) which includes a biologically active domain which is at least about 50% or more homologous to one of the amino acid sequences of Appendix B and is able to participate in the metabolism of compounds necessary for the construction of cellular membranes in *C glutamicum*, or in the transport of molecules across these membranes, or has one or more of the activities set forth in Table 1, and which also includes heterologous nucleic acid sequences encoding a heterologous polypeptide or regulatory regions.

In another embodiment, the isolated nucleic acid molecule is at least 15 nucleotides in length and hybridizes under stringent conditions to a nucleic acid molecule comprising a nucleotide sequence of Appendix A. Preferably, the isolated nucleic acid molecule corresponds to a naturally-occurring nucleic acid molecule. More preferably, the isolated nucleic acid encodes a naturally-occurring *C. glutamicum* MCT protein, or a biologically active portion thereof.

Another aspect of the invention pertains to vectors, e.g., recombinant expression vectors, containing the nucleic acid molecules of the invention, and host cells into which such vectors have been introduced. In one embodiment, such a host cell is used to produce an MCT protein by culturing the host cell in a suitable medium. The MCT protein can be then isolated from the medium or the host cell.

Yet another aspect of the invention pertains to a genetically altered microorganism in which an MCT gene has been introduced or altered. In one embodiment, the genome of the microorganism has been altered by introduction of a nucleic acid molecule of the invention encoding wild-type or mutated MCT sequence as a transgene. In another embodiment, an endogenous MCT gene within the genome of

the microorganism has been altered, e.g., functionally disrupted, by homologous recombination with an altered MCT gene. In a preferred embodiment, the microorganism belongs to the genus *Corynebacterium* or *Brevibacterium*, with *Corynebacterium glutamicum* being particularly preferred. In a preferred embodiment, the microorganism is also utilized for the production of a desired compound, such as an amino acid, with lysine being particularly preferred.

Still another aspect of the invention pertains to an isolated MCT protein or a portion, e.g., a biologically active portion, thereof. In a preferred embodiment, the isolated MCT protein or portion thereof can participate in the metabolism of compounds necessary for the construction of cellular membranes in *C. glutamicum*, or in the transport of molecules across these membranes. In another preferred embodiment, the isolated MCT protein or portion thereof is sufficiently homologous to an amino acid sequence of Appendix B such that the protein or portion thereof maintains the ability to participate in the metabolism of compounds necessary for the construction of cellular membranes in *C. glutamicum*, or in the transport of molecules across these membranes.

The invention also provides an isolated preparation of an MCT protein. In preferred embodiments, the MCT protein comprises an amino acid sequence of Appendix B. In another preferred embodiment, the invention pertains to an isolated full length protein which is substantially homologous to an entire amino acid sequence of Appendix B (encoded by an open reading frame set forth in Appendix A). In yet another embodiment, the protein is at least about 50%, preferably at least about 60%, and more preferably at least about 70%, 80%, or 90%, and most preferably at least about 95%, 96%, 97%, 98%, or 99% or more homologous to an entire amino acid sequence of Appendix B. In other embodiments, the isolated MCT protein comprises an amino acid sequence which is at least about 50% or more homologous to one of the amino acid sequences of Appendix B and is able to participate in the metabolism of compounds necessary for the construction of cellular membranes in *C. glutamicum*, or in the transport of molecules across these membranes, or has one or more of the activities set forth in Table 1.

Alternatively, the isolated MCT protein can comprise an amino acid sequence which is encoded by a nucleotide sequence which hybridizes, e.g., hybridizes under stringent conditions, or is at least about 50%, preferably at least about 60%, more preferably at least about 70%, 80%, or 90%, and even more preferably at least about 95%, 96%, 97%, 98,%, or 99% or more homologous, to a nucleotide sequence of Appendix B. It is also preferred that the preferred forms of MCT proteins also have one or more of the MCT bioactivities described herein.

The MCT polypeptide, or a biologically active portion thereof, can be operatively linked to a non-MCT polypeptide to form a fusion protein. In preferred embodiments, this fusion protein has an activity which differs from that of the MCT protein alone. In other preferred embodiments, this fusion protein participate in the metabolism of compounds necessary for the construction of cellular membranes in *C glutamicum*, or in the transport of molecules across these membranes. In particularly preferred embodiments, integration of this fusion protein into a host cell modulates production of a desired compound from the cell.

Another aspect of the invention pertains to a method for producing a fine chemical. This method involves the culturing of a cell containing a vector directing the expression of an MCT nucleic acid molecule of the invention, such that a fine chemical is produced. In a preferred embodiment, this method further includes the step of obtaining a cell containing such a vector, in which a cell is transfected with a vector directing the expression of an MCT nucleic acid. In another preferred embodiment, this method further includes the step of recovering the fine chemical from the culture. In a particularly preferred embodiment, the cell is from the genus *Corynebacterium* or *Brevibacterium*, or is selected from those strains set forth in Table 3.

Another aspect of the invention pertains to methods for modulating production of a molecule from a microorganism. Such methods include contacting the cell with an agent which modulates MCT protein activity or MCT nucleic acid expression such that a cell associated activity is altered relative to this same activity in the absence of the agent. In a preferred embodiment, the cell is modulated for one or more *C. glutamicum* metabolic pathways for cell membrane components or is modulated for the transport of compounds across such membranes, such that the yields or rate of production of a desired fine chemical by this microorganism is improved. The agent which modulates MCT protein activity can be an agent which stimulates MCT protein activity or MCT nucleic acid expression. Examples of agents which stimulate MCT protein activity or MCT nucleic acid expression include small molecules, active MCT proteins, and nucleic acids encoding MCT proteins that have been introduced into the cell. Examples of agents which inhibit MCT activity or expression include small molecules and antisense MCT nucleic acid molecules.

Another aspect of the invention pertains to methods for modulating yields of a desired compound from a cell, involving the introduction of a wild-type or mutant MCT gene into a cell, either maintained on a separate plasmid or integrated into the genome of the host cell. If integrated into the genome, such integration can be random, or it can take place by homologous recombination such that the native gene is replaced by the introduced copy, causing the production of the desired compound from the cell to be

modulated. In a preferred embodiment, said yields are increased. In another preferred embodiment, said chemical is a fine chemical. In a particularly preferred embodiment, said fine chemical is an amino acid. In especially preferred embodiments, said amino acid is L-lysine.

Detailed Description of the Invention

The present invention provides MCT nucleic acid and protein molecules which are involved in the metabolism of cellular membrane components in C. glutamicum or in the transport of compounds across such membranes. The molecules of the invention may be utilized in the modulation of production of fine chemicals from microorganisms, such as *C. glutamicum*, either directly (e.g., where overexpression or optimization of a fatty acid biosynthesis protein has a direct impact on the yield, production, and/or efficiency of production of the fatty acid from modified *C glutamicum*), or an indirect impact which nonetheless results in an increase of yield, production, and/or efficiency of production of the desired compound (e.g., where modulation of the metabolism of cell membrane components results in alterations in the yield, production, and/or efficiency of production or the composition of the cell membrane, which in turn may impact the production of one or more fine chemicals). Aspects of the invention are further explicated below.

I. Fine Chemicals

The term 'fine chemical' is art-recognized and includes molecules produced by an organism which have applications in various industries, such as, but not limited to, the pharmaceutical, agriculture, and cosmetics industries. Such compounds include organic acids, such as tartaric acid, itaconic acid, and diaminopimelic acid, both proteinogenic and non-proteinogenic amino acids, purine and pyrimidine bases, nucleosides, and nucleotides (as described e.g. in Kuninaka, A. (1996) Nucleotides and related compounds, p. 561-612, in Biotechnology vol. 6, Rehm et al., eds. VCH: Weinheim, and references contained therein), lipids, both saturated and unsaturated fatty acids (e.g., arachidonic acid), diols (e.g., propane diol, and butane diol), carbohydrates (e.g., hyaluronic acid and trehalose), aromatic compounds (e.g., aromatic amines, vanillin, and indigo), vitamins and cofactors (as described in Ullmann's Encyclopedia of Industrial Chemistry, vol. A27, "Vitamins", p. 443-613 (1996) VCH: Weinheim and references therein; and Ong, A.S., Niki, E. & Packer, L. (1995) "Nutrition, Lipids, Health, and Disease" Proceedings of the UNESCO/Confederation of Scientific and Technological Associations in Malaysia, and the Society for Free Radical Research – Asia, held Sept. 1-3, 1994 at Penang, Malaysia, AOCS Press, (1995)), enzymes, and all

other chemicals described in Gutcho (1983) Chemicals by Fermentation, Noyes Data Corporation, ISBN: 0818805086 and references therein. The metabolism and uses of certain of these fine chemicals are further explicated below.

A. *Amino Acid Metabolism and Uses*

Amino acids comprise the basic structural units of all proteins, and as such are essential for normal cellular functioning in all organisms. The term "amino acid" is art-recognized. The proteinogenic amino acids, of which there are 20 species, serve as structural units for proteins, in which they are linked by peptide bonds, while the nonproteinogenic amino acids (hundreds of which are known) are not normally found in proteins (see Ulmann's Encyclopedia of Industrial Chemistry, vol. A2, p. 57-97 VCH: Weinheim (1985)). Amino acids may be in the D- or L- optical configuration, though L-amino acids are generally the only type found in naturally-occurring proteins. Biosynthetic and degradative pathways of each of the 20 proteinogenic amino acids have been well characterized in both prokaryotic and eukaryotic cells (see, for example, Stryer, L. Biochemistry, 3rd edition, pages 578-590 (1988)). The 'essential' amino acids (histidine, isoleucine, leucine, lysine, methionine, phenylalanine, threonine, tryptophan, and valine), so named because they are generally a nutritional requirement due to the complexity of their biosynthesis, are readily converted by simple biosynthetic pathways to the remaining 11 'nonessential' amino acids (alanine, arginine, asparagine, aspartate, cysteine, glutamate, glutamine, glycine, proline, serine, and tyrosine). Higher animals do retain the ability to synthesize some of these amino acids, but the essential amino acids must be supplied from the diet in order for normal protein synthesis to occur.

Aside from their function in protein biosynthesis, these amino acids are interesting chemicals in their own right, and many have been found to have various applications in the food, feed, chemical, cosmetics, agriculture, and pharmaceutical industries. Lysine is an important amino acid in the nutrition not only of humans, but also of monogastric animals such as poultry and swine. Glutamate is most commonly used as a flavor additive (mono-sodium glutamate, MSG) and is widely used throughout the food industry, as are aspartate, phenylalanine, glycine, and cysteine. Glycine, L-methionine and tryptophan are all utilized in the pharmaceutical industry. Glutamine, valine, leucine, isoleucine, histidine, arginine, proline, serine and alanine are of use in both the pharmaceutical and cosmetics industries. Threonine, tryptophan, and D/ L-methionine are common feed additives. (Leuchtenberger, W. (1996) Amino aids – technical production and use, p. 466-502 in Rehm et al. (eds.) Biotechnology vol. 6, chapter 14a, VCH: Weinheim). Additionally, these amino acids have been found to be useful as precursors for the synthesis of synthetic amino acids and proteins, such as N-

acetylcysteine, S-carboxymethyl-L-cysteine, (S)-5-hydroxytryptophan, and others described in Ullmann's Encyclopedia of Industrial Chemistry, vol. A2, p. 57-97, VCH: Weinheim, 1985.

The biosynthesis of these natural amino acids in organisms capable of producing them, such as bacteria, has been well characterized (for review of bacterial amino acid biosynthesis and regulation thereof, see Umbarger, H.E.(1978) *Ann. Rev. Biochem.* 47: 533-606). Glutamate is synthesized by the reductive amination of α-ketoglutarate, an intermediate in the citric acid cycle. Glutamine, proline, and arginine are each subsequently produced from glutamate. The biosynthesis of serine is a three-step process beginning with 3-phosphoglycerate (an intermediate in glycolysis), and resulting in this amino acid after oxidation, transamination, and hydrolysis steps. Both cysteine and glycine are produced from serine; the former by the condensation of homocysteine with serine, and the latter by the transferal of the side-chain β-carbon atom to tetrahydrofolate, in a reaction catalyzed by serine transhydroxymethylase. Phenylalanine, and tyrosine are synthesized from the glycolytic and pentose phosphate pathway precursors erythrose 4-phosphate and phosphoenolpyruvate in a 9-step biosynthetic pathway that differ only at the final two steps after synthesis of prephenate. Tryptophan is also produced from these two initial molecules, but its synthesis is an 11-step pathway. Tyrosine may also be synthesized from phenylalanine, in a reaction catalyzed by phenylalanine hydroxylase. Alanine, valine, and leucine are all biosynthetic products of pyruvate, the final product of glycolysis. Aspartate is formed from oxaloacetate, an intermediate of the citric acid cycle. Asparagine, methionine, threonine, and lysine are each produced by the conversion of aspartate. Isoleucine is formed from threonine. A complex 9-step pathway results in the production of histidine from 5-phosphoribosyl-1-pyrophosphate, an activated sugar.

Amino acids in excess of the protein synthesis needs of the cell cannot be stored, and are instead degraded to provide intermediates for the major metabolic pathways of the cell (for review see Stryer, L. Biochemistry 3rd ed. Ch. 21 "Amino Acid Degradation and the Urea Cycle" p. 495-516 (1988)). Although the cell is able to convert unwanted amino acids into useful metabolic intermediates, amino acid production is costly in terms of energy, precursor molecules, and the enzymes necessary to synthesize them. Thus it is not surprising that amino acid biosynthesis is regulated by feedback inhibition, in which the presence of a particular amino acid serves to slow or entirely stop its own production (for overview of feedback mechanisms in amino acid biosynthetic pathways, see Stryer, L. Biochemistry, 3rd ed. Ch. 24: "Biosynthesis of Amino Acids and Heme" p. 575-600 (1988)). Thus, the output of any particular amino acid is limited by the amount of that amino acid present in the cell.

B. Vitamin, Cofactor, and Nutraceutical Metabolism and Uses

Vitamins, cofactors, and nutraceuticals comprise another group of molecules which the higher animals have lost the ability to synthesize and so must ingest, although they are readily synthesized by other organisms such as bacteria. These molecules are either bioactive substances themselves, or are precursors of biologically active substances which may serve as electron carriers or intermediates in a variety of metabolic pathways. Aside from their nutritive value, these compounds also have significant industrial value as coloring agents, antioxidants, and catalysts or other processing aids. (For an overview of the structure, activity, and industrial applications of these compounds, see, for example, Ullman's Encyclopedia of Industrial Chemistry, "Vitamins" vol. A27, p. 443-613. VCH: Weinheim, 1996.) The term "vitamin" is art-recognized, and includes nutrients which are required by an organism for normal functioning, but which that organism cannot synthesize by itself. The group of vitamins may encompass cofactors and nutraceutical compounds. The language "cofactor" includes nonproteinaceous compounds required for a normal enzymatic activity to occur. Such compounds may be organic or inorganic; the cofactor molecules of the invention are preferably organic. The term "nutraceutical" includes dietary supplements having health benefits in plants and animals, particularly humans. Examples of such molecules are vitamins, antioxidants, and also certain lipids (e.g., polyunsaturated fatty acids).

The biosynthesis of these molecules in organisms capable of producing them, such as bacteria, has been largely characterized (Ullman's Encyclopedia of Industrial Chemistry, "Vitamins" vol. A27, p. 443-613, VCH: Weinheim, 1996; Michal, G. (1999) Biochemical Pathways: An Atlas of Biochemistry and Molecular Biology, John Wiley & Sons; Ong, A.S., Niki, E. & Packer, L. (1995) "Nutrition, Lipids, Health, and Disease" Proceedings of the UNESCO/Confederation of Scientific and Technological Associations in Malaysia, and the Society for Free Radical Research – Asia, held Sept. 1-3, 1994 at Penang, Malaysia, AOCS Press: Champaign, IL X, 374 S).

Thiamin (vitamin B_1) is produced by the chemical coupling of pyrimidine and thiazole moieties. Riboflavin (vitamin B_2) is synthesized from guanosine-5'-triphosphate (GTP) and ribose-5'-phosphate. Riboflavin, in turn, is utilized for the synthesis of flavin mononucleotide (FMN) and flavin adenine dinucleotide (FAD). The family of compounds collectively termed 'vitamin B_6' (e.g., pyridoxine, pyridoxamine, pyridoxa-5'-phosphate, and the commercially used pyridoxin hydrochloride) are all derivatives of the common structural unit, 5-hydroxy-6-methylpyridine. Pantothenate (pantothenic acid, (R)-(+)-N-(2,4-dihydroxy-3,3-dimethyl-1-oxobutyl)-β-alanine) can be produced

either by chemical synthesis or by fermentation. The final steps in pantothenate biosynthesis consist of the ATP-driven condensation of β-alanine and pantoic acid. The enzymes responsible for the biosynthesis steps for the conversion to pantoic acid, to β-alanine and for the condensation to panthotenic acid are known. The metabolically active form of pantothenate is Coenzyme A, for which the biosynthesis proceeds in 5 enzymatic steps. Pantothenate, pyridoxal-5'-phosphate, cysteine and ATP are the precursors of Coenzyme A. These enzymes not only catalyze the formation of panthothante, but also the production of (R)-pantoic acid, (R)-pantolacton, (R)-panthenol (provitamin B_5), pantetheine (and its derivatives) and coenzyme A.

Biotin biosynthesis from the precursor molecule pimeloyl-CoA in microorganisms has been studied in detail and several of the genes involved have been identified. Many of the corresponding proteins have been found to also be involved in Fe-cluster synthesis and are members of the nifS class of proteins. Lipoic acid is derived from octanoic acid, and serves as a coenzyme in energy metabolism, where it becomes part of the pyruvate dehydrogenase complex and the α-ketoglutarate dehydrogenase complex. The folates are a group of substances which are all derivatives of folic acid, which is turn is derived from L-glutamic acid, p-amino-benzoic acid and 6-methylpterin. The biosynthesis of folic acid and its derivatives, starting from the metabolism intermediates guanosine-5'-triphosphate (GTP), L-glutamic acid and p-amino-benzoic acid has been studied in detail in certain microorganisms.

Corrinoids (such as the cobalamines and particularly vitamin B_{12}) and porphyrines belong to a group of chemicals characterized by a tetrapyrole ring system. The biosynthesis of vitamin B_{12} is sufficiently complex that it has not yet been completely characterized, but many of the enzymes and substrates involved are now known. Nicotinic acid (nicotinate), and nicotinamide are pyridine derivatives which are also termed 'niacin'. Niacin is the precursor of the important coenzymes NAD (nicotinamide adenine dinucleotide) and NADP (nicotinamide adenine dinucleotide phosphate) and their reduced forms.

The large-scale production of these compounds has largely relied on cell-free chemical syntheses, though some of these chemicals have also been produced by large-scale culture of microorganisms, such as riboflavin, Vitamin B_6, pantothenate, and biotin. Only Vitamin B_{12} is produced solely by fermentation, due to the complexity of its synthesis. *In vitro* methodologies require significant inputs of materials and time, often at great cost.

C. Purine, Pyrimidine, Nucleoside and Nucleotide Metabolism and Uses

Purine and pyrimidine metabolism genes and their corresponding proteins are important targets for the therapy of tumor diseases and viral infections. The language "purine" or "pyrimidine" includes the nitrogenous bases which are constituents of nucleic acids, co-enzymes, and nucleotides. The term "nucleotide" includes the basic structural units of nucleic acid molecules, which are comprised of a nitrogenous base, a pentose sugar (in the case of RNA, the sugar is ribose; in the case of DNA, the sugar is D-deoxyribose), and phosphoric acid. The language "nucleoside" includes molecules which serve as precursors to nucleotides, but which are lacking the phosphoric acid moiety that nucleotides possess. By inhibiting the biosynthesis of these molecules, or their mobilization to form nucleic acid molecules, it is possible to inhibit RNA and DNA synthesis; by inhibiting this activity in a fashion targeted to cancerous cells, the ability of tumor cells to divide and replicate may be inhibited. Additionally, there are nucleotides which do not form nucleic acid molecules, but rather serve as energy stores (i.e., AMP) or as coenzymes (i.e., FAD and NAD).

Several publications have described the use of these chemicals for these medical indications, by influencing purine and/or pyrimidine metabolism (e.g. Christopherson, R.I. and Lyons, S.D. (1990) "Potent inhibitors of *de novo* pyrimidine and purine biosynthesis as chemotherapeutic agents." *Med. Res. Reviews* 10: 505-548). Studies of enzymes involved in purine and pyrimidine metabolism have been focused on the development of new drugs which can be used, for example, as immunosuppressants or anti-proliferants (Smith, J.L., (1995) "Enzymes in nucleotide synthesis." *Curr. Opin. Struct. Biol.* 5: 752-757; (1995) *Biochem Soc. Transact.* 23: 877-902). However, purine and pyrimidine bases, nucleosides and nucleotides have other utilities: as intermediates in the biosynthesis of several fine chemicals (e.g., thiamine, S-adenosyl-methionine, folates, or riboflavin), as energy carriers for the cell (e.g., ATP or GTP), and for chemicals themselves, commonly used as flavor enhancers (e.g., IMP or GMP) or for several medicinal applications (see, for example, Kuninaka, A. (1996) Nucleotides and Related Compounds in Biotechnology vol. 6, Rehm et al., eds. VCH: Weinheim, p. 561-612). Also, enzymes involved in purine, pyrimidine, nucleoside, or nucleotide metabolism are increasingly serving as targets against which chemicals for crop protection, including fungicides, herbicides and insecticides, are developed.

The metabolism of these compounds in bacteria has been characterized (for reviews see, for example, Zalkin, H. and Dixon, J.E. (1992) "*de novo* purine nucleotide biosynthesis", in: Progress in Nucleic Acid Research and Molecular Biology, vol. 42, Academic Press:, p. 259-287; and Michal, G. (1999) "Nucleotides and Nucleosides", Chapter 8 in: Biochemical Pathways: An Atlas of Biochemistry and Molecular Biology,

Wiley: New York). Purine metabolism has been the subject of intensive research, and is essential to the normal functioning of the cell. Impaired purine metabolism in higher animals can cause severe disease, such as gout. Purine nucleotides are synthesized from ribose-5-phosphate, in a series of steps through the intermediate compound inosine-5'-phosphate (IMP), resulting in the production of guanosine-5'-monophosphate (GMP) or adenosine-5'-monophosphate (AMP), from which the triphosphate forms utilized as nucleotides are readily formed. These compounds are also utilized as energy stores, so their degradation provides energy for many different biochemical processes in the cell. Pyrimidine biosynthesis proceeds by the formation of uridine-5'-monophosphate (UMP) from ribose-5-phosphate. UMP, in turn, is converted to cytidine-5'-triphosphate (CTP). The deoxy- forms of all of these nucleotides are produced in a one step reduction reaction from the diphosphate ribose form of the nucleotide to the diphosphate deoxyribose form of the nucleotide. Upon phosphorylation, these molecules are able to participate in DNA synthesis.

D. Trehalose Metabolism and Uses

Trehalose consists of two glucose molecules, bound in α, α-1,1 linkage. It is commonly used in the food industry as a sweetener, an additive for dried or frozen foods, and in beverages. However, it also has applications in the pharmaceutical, cosmetics and biotechnology industries (see, for example, Nishimoto et al., (1998) U.S. Patent No. 5,759,610; Singer, M.A. and Lindquist, S. (1998) *Trends Biotech.* 16: 460-467; Paiva, C.L.A. and Panek, A.D. (1996) *Biotech Ann. Rev.* 2: 293-314; and Shiosaka, M. (1997) J. Japan 172: 97-102). Trehalose is produced by enzymes from many microorganisms and is naturally released into the surrounding medium, from which it can be collected using methods known in the art.

II. Membrane Biosynthesis and Transmembrane Transport

Cellular membranes serve a variety of functions in a cell. First and foremost, a membrane differentiates the contents of a cell from the surrounding environment, thus giving integrity to the cell. Membranes may also serve as barriers to the influx of hazardous or unwanted compounds, and also to the efflux of desired compounds. Cellular membranes are by nature impervious to the unfacilitated diffusion of hydrophilic compounds such as proteins, water molecules and ions due to their structure: a bilayer of lipid molecules in which the polar head groups face outwards (towards the exterior and interior of the cell, respectively) and the nonpolar tails face inwards at the center of the bilayer, forming a hydrophobic core (for a general review of membrane structure and function, see Gennis, R.B. (1989) Biomembranes, Molecular Structure and

Function, Springer: Heidelberg). This barrier enables cells to maintain a relatively higher concentration of desired compounds and a relatively lower concentration of undesired compounds than are contained within the surrounding medium, since the diffusion of these compounds is effectively blocked by the membrane.

However, the membrane also presents an effective barrier to the import of desired compounds and the export of waste molecules. To overcome this difficulty, cellular membranes incorporate many kinds of transporter proteins which are able to facilitate the transmembrane transport of different kinds of compounds. There are two general classes of these transport proteins: pores or channels and transporters. The former are integral membrane proteins, sometimes complexes of proteins, which form a regulated hole through the membrane. This regulation, or 'gating' is generally specific to the molecules to be transported by the pore or channel, rendering these transmembrane constructs selectively permeable to a specific class of substrates; for example, a potassium channel is constructed such that only ions having a like charge and size to that of potassium may pass through. Channel and pore proteins tend to have discrete hydrophobic and hydrophilic domains, such that the hydrophobic face of the protein may associate with the interior of the membrane while the hydrophilic face lines the interior of the channel, thus providing a sheltered hydrophilic environment through which the selected hydrophilic molecule may pass. Many such pores/channels are known in the art, including those for potassium, calcium, sodium, and chloride ions.

This pore and channel-mediated system of facilitated diffusion is limited to very small molecules, such as ions, because pores or channels large enough to permit the passage of whole proteins by facilitated diffusion would be unable to prevent the passage of smaller hydrophilic molecules as well. Transport of molecules by this process is sometimes termed 'facilitated diffusion' since the driving force of a concentration gradient is required for the transport to occur. Permeases also permit facilitated diffusion of larger molecules, such as glucose or other sugars, into the cell when the concentration of these molecules on one side of the membrane is greater than that on the other (also called 'uniport'). In contrast to pores or channels, these integral membrane proteins (often having between 6-14 membrane-spanning α-helices) do not form open channels through the membrane, but rather bind to the target molecule at the surface of the membrane and then undergo a conformational shift such that the target molecule is released on the opposite side of the membrane.

However, cells frequently require the import or export of molecules against the existing concentration gradient ('active transport'), a situation in which facilitated diffusion cannot occur. There are two general mechanisms used by cells for such membrane transport: symport or antiport, and energy-coupled transport such as that

mediated by the ABC transporters. Symport and antiport systems couple the movement of two different molecules across the membrane (via permeases having two separate binding sites for the two different molecules); in symport, both molecules are transported in the same direction, while in antiport, one molecule is imported while the other is exported. This is possible energetically because one of the two molecules moves in accordance with a concentration gradient, and this energetically favorable event is permitted only upon concomitant movement of a desired compound against the prevailing concentration gradient. Single molecules may be transported across the membrane against the concentration gradient in an energy-driven process, such as that utilized by the ABC transporters. In this system, the transport protein located in the membrane has an ATP-binding cassette; upon binding of the target molecule, the ATP is converted to ADP + Pi, and the resulting release of energy is used to drive the movement of the target molecule to the opposite face of the membrane, facilitated by the transporter. For more detailed descriptions of all of these transport systems, see: Bamberg, E. et al., (1993) "Charge transport of ion pumps on lipid bilayer membranes", *Q. Rev. Biophys.* 26: 1-25; Findlay, J.B.C. (1991) "Structure and function in membrane transport systems", *Curr Opin. Struct. Biol.* 1:804-810; Higgins, C.F. (1992) "ABC transporters from microorganisms to man", *Ann. Rev. Cell Biol.* 8: 67-113; Gennis, R.B. (1989) "Pores, Channels and Transporters", in: Biomembranes, Molecular Structure and Function, Springer: Heidelberg, p. 270-322; and Nikaido, H. and Saier, H. (1992) "Transport proteins in bacteria: common themes in their design", *Science* 258: 936-942, and references contained within each of these references.

The synthesis of membranes is a well-characterized process involving a number of components, the most important of which are lipid molecules. Lipid synthesis may be divided into two parts: the synthesis of fatty acids and their attachment to sn-glycerol-3-phosphate, and the addition or modification of a polar head group. Typical lipids utilized in bacterial membranes include phospholipids, glycolipids, sphingolipids, and phosphoglycerides. Fatty acid synthesis begins with the conversion of acetyl CoA either to malonyl CoA by acetyl CoA carboxylase, or to acetyl-ACP by acetyltransacylase. Following a condensation reaction, these two product molecules together form acetoacetyl-ACP, which is converted by a series of condensation, reduction and dehydration reactions to yield a saturated fatty acid molecule having a desired chain length. The production of unsaturated fatty acids from such molecules is catalyzed by specific desaturases either aerobically, with the help of molecular oxygen, or anaerobically (for reference on fatty acid synthesis, see F.C. Neidhardt et al. (1996) E. coli and Salmonella. ASM Press: Washington, D.C., p. 612-636 and references contained therein; Lengeler et al. (eds) (1999) Biology of Procaryotes. Thieme:

Stuttgart, New York, and references contained therein; and Magnuson, K. et al., (1993) *Microbiological Reviews* 57: 522-542, and references contained therein). The cyclopropane fatty acids (CFA) are synthesized by a specific CFA-synthase using SAM as a cosubstrate. Branched chain fatty acids are synthesized from branched chain amino acids that are deaminated to yield branched chain 2-oxo-acids (see Lengeler et al., eds. (1999) Biology of Procaryotes. Thieme: Stuttgart, New York, and references contained therein). Another essential step in lipid synthesis is the transfer of fatty acids onto the polar head groups by, for example, glycerol-phosphate-acyltransferases. The combination of various precursor molecules and biosynthetic enzymes results in the production of different fatty acid molecules, which has a profound effect on the composition of the membrane.

III. Elements and Methods of the Invention

The present invention is based, at least in part, on the discovery of novel molecules, referred to herein as MCT nucleic acid and protein molecules, which control the production of cellular membranes in *C. glutamicum* and govern the movement of molecules across such membranes. In one embodiment, the MCT molecules participate in the metabolism of compounds necessary for the construction of cellular membranes in *C. glutamicum*, or in the transport of molecules across these membranes. In a preferred embodiment, the activity of the MCT molecules of the present invention to regulate membrane component production and membrane transport has an impact on the production of a desired fine chemical by this organism. In a particularly preferred embodiment, the MCT molecules of the invention are modulated in activity, such that the *C. glutamicum* metabolic pathways which the MCT proteins of the invention regulate are modulated in yield, production, and/or efficiency of production and the transport of compounds through the membranes is altered in efficiency, which either directly or indirectly modulates the yield, production, and/or efficiency of production of a desired fine chemical by *C. glutamicum*.

The language, "MCT protein" or "MCT polypeptide" includes proteins which participate in the metabolism of compounds necessary for the construction of cellular membranes in *C. glutamicum*, or in the transport of molecules across these membranes. Examples of MCT proteins include those encoded by the MCT genes set forth in Table 1 and Appendix A. The terms "MCT gene" or "MCT nucleic acid sequence" include nucleic acid sequences encoding an MCT protein, which consist of a coding region and also corresponding untranslated 5' and 3' sequence regions. Examples of MCT genes include those set forth in Table 1. The terms "production" or "productivity" are art-recognized and include the concentration of the fermentation product (for example, the

desired fine chemical) formed within a given time and a given fermentation volume (e.g., kg product per hour per liter). The term "efficiency of production" includes the time required for a particular level of production to be achieved (for example, how long it takes for the cell to attain a particular rate of output of a fine chemical). The term "yield" or "product/carbon yield" is art-recognized and includes the efficiency of the conversion of the carbon source into the product (i.e., fine chemical). This is generally written as, for example, kg product per kg carbon source. By increasing the yield or production of the compound, the quantity of recovered molecules, or of useful recovered molecules of that compound in a given amount of culture over a given amount of time is increased. The terms "biosynthesis" or a "biosynthetic pathway" are art-recognized and include the synthesis of a compound, preferably an organic compound, by a cell from intermediate compounds in what may be a multistep and highly regulated process. The terms "degradation" or a "degradation pathway" are art-recognized and include the breakdown of a compound, preferably an organic compound, by a cell to degradation products (generally speaking, smaller or less complex molecules) in what may be a multistep and highly regulated process. The language "metabolism" is art-recognized and includes the totality of the biochemical reactions that take place in an organism. The metabolism of a particular compound, then, (e.g., the metabolism of an amino acid such as glycine) comprises the overall biosynthetic, modification, and degradation pathways in the cell related to this compound.

In another embodiment, the MCT molecules of the invention are capable of modulating the production of a desired molecule, such as a fine chemical, in a microorganism such as *C. glutamicum*. There are a number of mechanisms by which the alteration of an MCT protein of the invention may directly affect the yield, production, and/or efficiency of production of a fine chemical from a *C. glutamicum* strain incorporating such an altered protein. Those MCT proteins involved in the export of fine chemical molecules from the cell may be increased in number or activity such that greater quantities of these compounds are secreted to the extracellular medium, from which they are more readily recovered. Similarly, those MCT proteins involved in the import of nutrients necessary for the biosynthesis of one or more fine chemicals (e.g., phosphate, sulfate, nitrogen compounds, etc.) may be increased in number or activity such that these precursor , cofactor, or intermediate compounds are increased in concentration within the cell. Further, fatty acids and lipids themselves are desirable fine chemicals; by optimizing the activity or increasing the number of one or more MCT proteins of the invention which participate in the biosynthesis of these compounds, or by impairing the activity of one or more MCT proteins which are involved in the

degradation of these compounds, it may be possible to increase the yield, production, and/or efficiency of production of fatty acid and lipid molecules from *C. glutamicum*.

The mutagenesis of one or more MCT genes of the invention may also result in MCT proteins having altered activities which indirectly impact the production of one or more desired fine chemicals from *C .glutamicum*. For example, MCT proteins of the invention involved in the export of waste products may be increased in number or activity such that the normal metabolic wastes of the cell (possibly increased in quantity due to the overproduction of the desired fine chemical) are efficiently exported before they are able to damage nucleotides and proteins within the cell (which would decrease the viability of the cell) or to interfere with fine chemical biosynthetic pathways (which would decrease the yield, production, or efficiency of production of the desired fine chemical). Further, the relatively large intracellular quantities of the desired fine chemical may in itself be toxic to the cell, so by increasing the activity or number of transporters able to export this compound from the cell, one may increase the viability of the cell in culture, in turn leading to a greater number of cells in the culture producing the desired fine chemical. The MCT proteins of the invention may also be manipulated such that the relative amounts of different lipid and fatty acid molecules are produced. This may have a profound effect on the lipid composition of the membrane of the cell. Since each type of lipid has different physical properties, an alteration in the lipid composition of a membrane may significantly alter membrane fluidity. Changes in membrane fluidity can impact the transport of molecules across the membrane, as well as the integrity of the cell, both of which have a profound effect on the production of fine chemicals from *C. glutamicum* in large-scale fermentative culture.

The isolated nucleic acid sequences of the invention are contained within the genome of a *Corynebacterium glutamicum* strain available through the American Type Culture Collection, given designation ATCC 13032. The nucleotide sequence of the isolated *C. glutamicum* MCT cDNAs and the predicted amino acid sequences of the *C. glutamicum* MCT proteins are shown in Appendices A and B, respectively. Computational analyses were performed which classified and/or identified these nucleotide sequences as sequences which encode proteins involved in the metabolism of cellular membrane components or proteins involved in the transport of compounds across such membranes.

The present invention also pertains to proteins which have an amino acid sequence which is substantially homologous to an amino acid sequence of Appendix B. As used herein, a protein which has an amino acid sequence which is substantially homologous to a selected amino acid sequence is least about 50% homologous to the selected amino acid sequence, e.g., the entire selected amino acid sequence. A protein

which has an amino acid sequence which is substantially homologous to a selected amino acid sequence can also be least about 50-60%, preferably at least about 60-70%, and more preferably at least about 70-80%, 80-90%, or 90-95%, and most preferably at least about 96%, 97%, 98%, 99% or more homologous to the selected amino acid sequence.

The MCT protein or a biologically active portion or fragment thereof of the invention can participate in the metabolism of compounds necessary for the construction of cellular membranes in *C. glutamicum*, or in the transport of molecules across these membranes, or have one or more of the activities set forth in Table 1.

Various aspects of the invention are described in further detail in the following subsections:

A. Isolated Nucleic Acid Molecules

One aspect of the invention pertains to isolated nucleic acid molecules that encode MCT polypeptides or biologically active portions thereof, as well as nucleic acid fragments sufficient for use as hybridization probes or primers for the identification or amplification of MCT-encoding nucleic acid (e.g., MCT DNA). As used herein, the term "nucleic acid molecule" is intended to include DNA molecules (e.g., cDNA or genomic DNA) and RNA molecules (e.g., mRNA) and analogs of the DNA or RNA generated using nucleotide analogs. This term also encompasses untranslated sequence located at both the 3' and 5' ends of the coding region of the gene: at least about 100 nucleotides of sequence upstream from the 5' end of the coding region and at least about 20 nucleotides of sequence downstream from the 3'end of the coding region of the gene. The nucleic acid molecule can be single-stranded or double-stranded, but preferably is double-stranded DNA. An "isolated" nucleic acid molecule is one which is separated from other nucleic acid molecules which are present in the natural source of the nucleic acid. Preferably, an "isolated" nucleic acid is free of sequences which naturally flank the nucleic acid (i.e., sequences located at the 5' and 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated MCT nucleic acid molecule can contain less than about 5 kb, 4kb, 3kb, 2kb, 1 kb, 0.5 kb or 0.1 kb of nucleotide sequences which naturally flank the nucleic acid molecule in genomic DNA of the cell from which the nucleic acid is derived (e.g, a *C glutamicum* cell). Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material, or culture medium when produced by recombinant techniques, or chemical precursors or other chemicals when chemically synthesized.

A nucleic acid molecule of the present invention, e.g., a nucleic acid molecule having a nucleotide sequence of Appendix A, or a portion thereof, can be isolated using standard molecular biology techniques and the sequence information provided herein. For example, a *C. glutamicum* MCT cDNA can be isolated from a *C. glutamicum* library using all or portion of one of the sequences of Appendix A as a hybridization probe and standard hybridization techniques (e.g., as described in Sambrook, J., Fritsh, E. F., and Maniatis, T. *Molecular Cloning: A Laboratory Manual. 2nd. ed., Cold Spring Harbor Laboratory*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989). Moreover, a nucleic acid molecule encompassing all or a portion of one of the sequences of Appendix A can be isolated by the polymerase chain reaction using oligonucleotide primers designed based upon this sequence (e.g., a nucleic acid molecule encompassing all or a portion of one of the sequences of Appendix A can be isolated by the polymerase chain reaction using oligonucleotide primers designed based upon this same sequence of Appendix A). For example, mRNA can be isolated from normal endothelial cells (e.g., by the guanidinium-thiocyanate extraction procedure of Chirgwin et al. (1979) *Biochemistry* 18: 5294-5299) and cDNA can be prepared using reverse transcriptase (e.g., Moloney MLV reverse transcriptase, available from Gibco/BRL, Bethesda, MD; or AMV reverse transcriptase, available from Seikagaku America, Inc., St. Petersburg, FL). Synthetic oligonucleotide primers for polymerase chain reaction amplification can be designed based upon one of the nucleotide sequences shown in Appendix A. A nucleic acid of the invention can be amplified using cDNA or, alternatively, genomic DNA, as a template and appropriate oligonucleotide primers according to standard PCR amplification techniques. The nucleic acid so amplified can be cloned into an appropriate vector and characterized by DNA sequence analysis. Furthermore, oligonucleotides corresponding to an MCT nucleotide sequence can be prepared by standard synthetic techniques, e.g., using an automated DNA synthesizer.

In a preferred embodiment, an isolated nucleic acid molecule of the invention comprises one of the nucleotide sequences shown in Appendix A. The sequences of Appendix A correspond to the *Corynebacterium glutamicum* MCT cDNAs of the invention. This cDNA comprises sequences encoding MCT proteins (i.e., "the coding region", indicated in each sequence in Appendix A), as well as 5' untranslated sequences and 3' untranslated sequences, also indicated in Appendix A. Alternatively, the nucleic acid molecule can comprise only the coding region of any of the sequences in Appendix A.

For the purposes of this application, it will be understood that each of the sequences set forth in Appendix A has an identifying RXA number having the designation "RXA" followed by 5 digits (i.e., RXA00001). Each of these sequences

comprises up to three parts: a 5' upstream region, a coding region, and a downstream region. Each of these three regions is identified by the same RXA designation to eliminate confusion. The recitation "one of the sequences in Appendix A", then, refers to any of the sequences in Appendix A, which may be distinguished by their differing RXA designations. The coding region of each of these sequences is translated into a corresponding amino acid sequence, which is set forth in Appendix B. The sequences of Appendix B are identified by the same RXA designations as Appendix A, such that they can be readily correlated. For example, the amino acid sequence in Appendix B designated RXA00001 is a translation of the coding region of the nucleotide sequence of nucleic acid molecule RXA00001 in Appendix A.

In one embodiment, the nucleic acid molecules of the present invention are not intended to include those compiled in Table 2. In the case of the dapD gene, a sequence for this gene was published in Wehrmann, A., et al. (1998) *J. Bacteriol.* 180(12): 3159-3165. However, the sequence obtained by the inventors of the present application is significantly longer than the published version. It is believed that the published version relied on an incorrect start codon, and thus represents only a fragment of the actual coding region.

In another preferred embodiment, an isolated nucleic acid molecule of the invention comprises a nucleic acid molecule which is a complement of one of the nucleotide sequences shown in Appendix A, or a portion thereof. A nucleic acid molecule which is complementary to one of the nucleotide sequences shown in Appendix A is one which is sufficiently complementary to one of the nucleotide sequences shown in Appendix A such that it can hybridize to one of the nucleotide sequences shown in Appendix A, thereby forming a stable duplex.

In still another preferred embodiment, an isolated nucleic acid molecule of the invention comprises a nucleotide sequence which is at least about 50-60%, preferably at least about 60-70%, more preferably at least about 70-80%, 80-90%, or 90-95%, and even more preferably at least about 95%, 96%, 97%, 98%, 99% or more homologous to a nucleotide sequence shown in Appendix A, or a portion thereof. In an additional preferred embodiment, an isolated nucleic acid molecule of the invention comprises a nucleotide sequence which hybridizes, e.g., hybridizes under stringent conditions, to one of the nucleotide sequences shown in Appendix A, or a portion thereof.

Moreover, the nucleic acid molecule of the invention can comprise only a portion of the coding region of one of the sequences in Appendix A, for example a fragment which can be used as a probe or primer or a fragment encoding a biologically active portion of an MCT protein. The nucleotide sequences determined from the cloning of the MCT genes from *C. glutamicum* allows for the generation of probes and

primers designed for use in identifying and/or cloning MCT homologues in other cell types and organisms, as well as MCT homologues from other *Corynebacteria* or related species. The probe/primer typically comprises substantially purified oligonucleotide. The oligonucleotide typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 12, preferably about 25, more preferably about 40, 50 or 75 consecutive nucleotides of a sense strand of one of the sequences set forth in Appendix A, an anti-sense sequence of one of the sequences set forth in Appendix A, or naturally occurring mutants thereof. Primers based on a nucleotide sequence of Appendix A can be used in PCR reactions to clone MCT homologues. Probes based on the MCT nucleotide sequences can be used to detect transcripts or genomic sequences encoding the same or homologous proteins. In preferred embodiments, the probe further comprises a label group attached thereto, e.g. the label group can be a radioisotope, a fluorescent compound, an enzyme, or an enzyme co-factor. Such probes can be used as a part of a diagnostic test kit for identifying cells which misexpress an MCT protein, such as by measuring a level of an MCT-encoding nucleic acid in a sample of cells, e.g., detecting MCT mRNA levels or determining whether a genomic MCT gene has been mutated or deleted.

In one embodiment, the nucleic acid molecule of the invention encodes a protein or portion thereof which includes an amino acid sequence which is sufficiently homologous to an amino acid sequence of Appendix B such that the protein or portion thereof maintains the ability to participate in the metabolism of compounds necessary for the construction of cellular membranes in *C. glutamicum*, or in the transport of molecules across these membranes. As used herein, the language "sufficiently homologous" refers to proteins or portions thereof which have amino acid sequences which include a minimum number of identical or equivalent (e.g., an amino acid residue which has a similar side chain as an amino acid residue in one of the sequences of Appendix B) amino acid residues to an amino acid sequence of Appendix B such that the protein or portion thereof is able to participate in the metabolism of compounds necessary for the construction of cellular membranes in *C. glutamicum*, or in the transport of molecules across these membranes. Protein members of such membrane component metabolic pathways or membrane transport systems, as described herein, may play a role in the production and secretion of one or more fine chemicals. Examples of such activities are also described herein. Thus, "the function of an MCT protein" contributes either directly or indirectly to the yield, production, and/or efficiency of production of one or more fine chemicals. Examples of MCT protein activities are set forth in Table 1.

In another embodiment, the protein is at least about 50-60%, preferably at least about 60-70%, and more preferably at least about 70-80%, 80-90%, 90-95%, and most preferably at least about 96%, 97%, 98%, 99% or more homologous to an entire amino acid sequence of Appendix B.

Portions of proteins encoded by the MCT nucleic acid molecules of the invention are preferably biologically active portions of one of the MCT proteins. As used herein, the term "biologically active portion of an MCT protein" is intended to include a portion, e.g., a domain/motif, of an MCT protein that participates in the metabolism of compounds necessary for the construction of cellular membranes in *C. glutamicum*, or in the transport of molecules across these membranes, or has an activity as set forth in Table 1. To determine whether an MCT protein or a biologically active portion thereof can participate in the metabolism of compounds necessary for the construction of cellular membranes in *C. glutamicum*, or in the transport of molecules across these membranes, an assay of enzymatic activity may be performed. Such assay methods are well known to those skilled in the art, as detailed in Example 8 of the Exemplification.

Additional nucleic acid fragments encoding biologically active portions of an MCT protein can be prepared by isolating a portion of one of the sequences in Appendix B, expressing the encoded portion of the MCT protein or peptide (e.g., by recombinant expression *in vitro*) and assessing the activity of the encoded portion of the MCT protein or peptide.

The invention further encompasses nucleic acid molecules that differ from one of the nucleotide sequences shown in Appendix A (and portions thereof) due to degeneracy of the genetic code and thus encode the same MCT protein as that encoded by the nucleotide sequences shown in Appendix A. In another embodiment, an isolated nucleic acid molecule of the invention has a nucleotide sequence encoding a protein having an amino acid sequence shown in Appendix B. In a still further embodiment, the nucleic acid molecule of the invention encodes a full length *C. glutamicum* protein which is substantially homologous to an amino acid sequence of Appendix B (encoded by an open reading frame shown in Appendix A).

In addition to the *C. glutamicum* MCT nucleotide sequences shown in Appendix A, it will be appreciated by those skilled in the art that DNA sequence polymorphisms that lead to changes in the amino acid sequences of MCT proteins may exist within a population (e.g., the *C. glutamicum* population). Such genetic polymorphism in the MCT gene may exist among individuals within a population due to natural variation. As used herein, the terms "gene" and "recombinant gene" refer to nucleic acid molecules comprising an open reading frame encoding an MCT protein, preferably a *C. glutamicum* MCT protein. Such natural variations can typically result in 1-5% variance

in the nucleotide sequence of the MCT gene. Any and all such nucleotide variations and resulting amino acid polymorphisms in MCT that are the result of natural variation and that do not alter the functional activity of MCT proteins are intended to be within the scope of the invention.

Nucleic acid molecules corresponding to natural variants and non-*C. glutamicum* homologues of the *C. glutamicum* MCT cDNA of the invention can be isolated based on their homology to the *C. glutamicum* MCT nucleic acid disclosed herein using the *C. glutamicum* cDNA, or a portion thereof, as a hybridization probe according to standard hybridization techniques under stringent hybridization conditions. Accordingly, in another embodiment, an isolated nucleic acid molecule of the invention is at least 15 nucleotides in length and hybridizes under stringent conditions to the nucleic acid molecule comprising a nucleotide sequence of Appendix A. In other embodiments, the nucleic acid is at least 30, 50, 100, 250 or more nucleotides in length. As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences at least 60% homologous to each other typically remain hybridized to each other. Preferably, the conditions are such that sequences at least about 65%, more preferably at least about 70%, and even more preferably at least about 75% or more homologous to each other typically remain hybridized to each other. Such stringent conditions are known to those skilled in the art and can be found in *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. A preferred, non-limiting example of stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2 X SSC, 0.1% SDS at 50-65°C. Preferably, an isolated nucleic acid molecule of the invention that hybridizes under stringent conditions to a sequence of Appendix A corresponds to a naturally-occurring nucleic acid molecule. As used herein, a "naturally-occurring" nucleic acid molecule refers to an RNA or DNA molecule having a nucleotide sequence that occurs in nature (e.g., encodes a natural protein). In one embodiment, the nucleic acid encodes a natural *C. glutamicum* MCT protein.

In addition to naturally-occurring variants of the MCT sequence that may exist in the population, the skilled artisan will further appreciate that changes can be introduced by mutation into a nucleotide sequence of Appendix A, thereby leading to changes in the amino acid sequence of the encoded MCT protein, without altering the functional ability of the MCT protein. For example, nucleotide substitutions leading to amino acid substitutions at "non-essential" amino acid residues can be made in a sequence of Appendix A. A "non-essential" amino acid residue is a residue that can be altered from the wild-type sequence of one of the MCT proteins (Appendix B) without altering the

activity of said MCT protein, whereas an "essential" amino acid residue is required for MCT protein activity. Other amino acid residues, however, (e.g., those that are not conserved or only semi-conserved in the domain having MCT activity) may not be essential for activity and thus are likely to be amenable to alteration without altering MCT activity.

Accordingly, another aspect of the invention pertains to nucleic acid molecules encoding MCT proteins that contain changes in amino acid residues that are not essential for MCT activity. Such MCT proteins differ in amino acid sequence from a sequence contained in Appendix B yet retain at least one of the MCT activities described herein. In one embodiment, the isolated nucleic acid molecule comprises a nucleotide sequence encoding a protein, wherein the protein comprises an amino acid sequence at least about 50% homologous to an amino acid sequence of Appendix B and is capable of participate in the metabolism of compounds necessary for the construction of cellular membranes in *C. glutamicum*, or in the transport of molecules across these membranes, or has one or more activities set forth in Table 1. Preferably, the protein encoded by the nucleic acid molecule is at least about 50-60% homologous to one of the sequences in Appendix B, more preferably at least about 60-70% homologous to one of the sequences in Appendix B, even more preferably at least about 70-80%, 80-90%, 90-95% homologous to one of the sequences in Appendix B, and most preferably at least about 96%, 97%, 98%, or 99% homologous to one of the sequences in Appendix B.

To determine the percent homology of two amino acid sequences (e.g., one of the sequences of Appendix B and a mutant form thereof) or of two nucleic acids, the sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in the sequence of one protein or nucleic acid for optimal alignment with the other protein or nucleic acid). The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in one sequence (e.g., one of the sequences of Appendix B) is occupied by the same amino acid residue or nucleotide as the corresponding position in the other sequence (e.g., a mutant form of the sequence selected from Appendix B), then the molecules are homologous at that position (i.e., as used herein amino acid or nucleic acid "homology" is equivalent to amino acid or nucleic acid "identity"). The percent homology between the two sequences is a function of the number of identical positions shared by the sequences (i.e., % homology = # of identical positions/total # of positions x 100).

An isolated nucleic acid molecule encoding an MCT protein homologous to a protein sequence of Appendix B can be created by introducing one or more nucleotide substitutions, additions or deletions into a nucleotide sequence of Appendix A such that one or more amino acid substitutions, additions or deletions are introduced into the

encoded protein. Mutations can be introduced into one of the sequences of Appendix A by standard techniques, such as site-directed mutagenesis and PCR-mediated mutagenesis. Preferably, conservative amino acid substitutions are made at one or more predicted non-essential amino acid residues. A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined in the art. These families include amino acids with basic side chains (e.g., lysine, arginine, histidine), acidic side chains (e.g., aspartic acid, glutamic acid), uncharged polar side chains (e.g., glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (e.g., alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (e.g., threonine, valine, isoleucine) and aromatic side chains (e.g., tyrosine, phenylalanine, tryptophan, histidine). Thus, a predicted nonessential amino acid residue in an MCT protein is preferably replaced with another amino acid residue from the same side chain family. Alternatively, in another embodiment, mutations can be introduced randomly along all or part of an MCT coding sequence, such as by saturation mutagenesis, and the resultant mutants can be screened for an MCT activity described herein to identify mutants that retain MCT activity. Following mutagenesis of one of the sequences of Appendix A, the encoded protein can be expressed recombinantly and the activity of the protein can be determined using, for example, assays described herein (see Example 8 of the Exemplification).

In addition to the nucleic acid molecules encoding MCT proteins described above, another aspect of the invention pertains to isolated nucleic acid molecules which are antisense thereto. An "antisense" nucleic acid comprises a nucleotide sequence which is complementary to a "sense" nucleic acid encoding a protein, e.g., complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence. Accordingly, an antisense nucleic acid can hydrogen bond to a sense nucleic acid. The antisense nucleic acid can be complementary to an entire MCT coding strand, or to only a portion thereof. In one embodiment, an antisense nucleic acid molecule is antisense to a "coding region" of the coding strand of a nucleotide sequence encoding an MCT protein. The term "coding region" refers to the region of the nucleotide sequence comprising codons which are translated into amino acid residues (e.g., the entire coding region of SEQ ID RXA00001 comprises nucleotides 1 to 1128). In another embodiment, the antisense nucleic acid molecule is antisense to a "noncoding region" of the coding strand of a nucleotide sequence encoding MCT. The term "noncoding region" refers to 5' and 3' sequences

which flank the coding region that are not translated into amino acids (i.e., also referred to as 5' and 3' untranslated regions).

Given the coding strand sequences encoding MCT disclosed herein (e.g., the sequences set forth in Appendix A), antisense nucleic acids of the invention can be designed according to the rules of Watson and Crick base pairing. The antisense nucleic acid molecule can be complementary to the entire coding region of MCT mRNA, but more preferably is an oligonucleotide which is antisense to only a portion of the coding or noncoding region of MCT mRNA. For example, the antisense oligonucleotide can be complementary to the region surrounding the translation start site of MCT mRNA. An antisense oligonucleotide can be, for example, about 5, 10, 15, 20, 25, 30, 35, 40, 45 or 50 nucleotides in length. An antisense nucleic acid of the invention can be constructed using chemical synthesis and enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (e.g., an antisense oligonucleotide) can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, e.g., phosphorothioate derivatives and acridine substituted nucleotides can be used. Examples of modified nucleotides which can be used to generate the antisense nucleic acid include 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-(carboxyhydroxylmethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine. Alternatively, the antisense nucleic acid can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (i.e., RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

The antisense nucleic acid molecules of the invention are typically administered to a cell or generated *in situ* such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding an MCT protein to thereby inhibit expression of the

protein, e.g., by inhibiting transcription and/or translation. The hybridization can be by conventional nucleotide complementarity to form a stable duplex, or, for example, in the case of an antisense nucleic acid molecule which binds to DNA duplexes, through specific interactions in the major groove of the double helix. The antisense molecule can be modified such that it specifically binds to a receptor or an antigen expressed on a selected cell surface, e.g., by linking the antisense nucleic acid molecule to a peptide or an antibody which binds to a cell surface receptor or antigen. The antisense nucleic acid molecule can also be delivered to cells using the vectors described herein. To achieve sufficient intracellular concentrations of the antisense molecules, vector constructs in which the antisense nucleic acid molecule is placed under the control of a strong pol II or pol III promoter are preferred.

In yet another embodiment, the antisense nucleic acid molecule of the invention is an α-anomeric nucleic acid molecule. An α-anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual β-units, the strands run parallel to each other (Gaultier et al. (1987) *Nucleic Acids. Res.* 15:6625-6641). The antisense nucleic acid molecule can also comprise a 2'-o-methylribonucleotide (Inoue et al. (1987) *Nucleic Acids Res.* 15:6131-6148) or a chimeric RNA-DNA analogue (Inoue et al. (1987) *FEBS Lett* 215:327-330).

In still another embodiment, an antisense nucleic acid of the invention is a ribozyme. Ribozymes are catalytic RNA molecules with ribonuclease activity which are capable of cleaving a single-stranded nucleic acid, such as an mRNA, to which they have a complementary region. Thus, ribozymes (e.g., hammerhead ribozymes (described in Haselhoff and Gerlach (1988) *Nature* 334:585-591)) can be used to catalytically cleave MCT mRNA transcripts to thereby inhibit translation of MCT mRNA. A ribozyme having specificity for an MCT-encoding nucleic acid can be designed based upon the nucleotide sequence of an MCT cDNA disclosed herein (i.e., RXA00001 in Appendix A). For example, a derivative of a *Tetrahymena* L-19 IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in an MCT-encoding mRNA. See, e.g., Cech et al. U.S. Patent No. 4,987,071 and Cech et al. U.S. Patent No. 5,116,742. Alternatively, MCT mRNA can be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules. See, e.g., Bartel, D. and Szostak, J.W. (1993) *Science* 261:1411-1418.

Alternatively, MCT gene expression can be inhibited by targeting nucleotide sequences complementary to the regulatory region of an MCT nucleotide sequence (e.g., an MCT promoter and/or enhancers) to form triple helical structures that prevent transcription of an MCT gene in target cells. See generally, Helene, C. (1991)

Anticancer Drug Des. 6(6):569-84; Helene, C. et al. (1992) *Ann. N.Y. Acad. Sci.* 660:27-36; and Maher, L.J. (1992) *Bioassays* 14(12):807-15.

B. Recombinant Expression Vectors and Host Cells

Another aspect of the invention pertains to vectors, preferably expression vectors, containing a nucleic acid encoding an MCT protein (or a portion thereof). As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded DNA loop into which additional DNA segments can be ligated. Another type of vector is a viral vector, wherein additional DNA segments can be ligated into the viral genome. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (e.g., bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (e.g., non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively linked. Such vectors are referred to herein as "expression vectors". In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. In the present specification, "plasmid" and "vector" can be used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include such other forms of expression vectors, such as viral vectors (e.g., replication defective retroviruses, adenoviruses and adeno-associated viruses), which serve equivalent functions.

The recombinant expression vectors of the invention comprise a nucleic acid of the invention in a form suitable for expression of the nucleic acid in a host cell, which means that the recombinant expression vectors include one or more regulatory sequences, selected on the basis of the host cells to be used for expression, which is operatively linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory sequence(s) in a manner which allows for expression of the nucleotide sequence (e.g., in an *in vitro* transcription/translation system or in a host cell when the vector is introduced into the host cell). The term "regulatory sequence" is intended to include promoters, enhancers and other expression control elements (e.g., polyadenylation signals). Such regulatory sequences are described, for example, in Goeddel; *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, CA (1990). Regulatory sequences include those which direct constitutive expression of a nucleotide sequence in many types of host cell and

those which direct expression of the nucleotide sequence only in certain host cells. It will be appreciated by those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, etc. The expression vectors of the invention can be introduced into host cells to thereby produce proteins or peptides, including fusion proteins or peptides, encoded by nucleic acids as described herein (e.g., MCT proteins, mutant forms of MCT proteins, fusion proteins, etc.).

The recombinant expression vectors of the invention can be designed for expression of MCT proteins in prokaryotic or eukaryotic cells. For example, MCT genes can be expressed in bacterial cells such as *C. glutamicum*, insect cells (using baculovirus expression vectors), yeast and other fungal cells (see Romanos, M.A. et al. (1992) "Foreign gene expression in yeast: a review", *Yeast* 8: 423-488; van den Hondel, C.A.M.J.J. et al. (1991) "Heterologous gene expression in filamentous fungi" in: More Gene Manipulations in Fungi, J.W. Bennet & L.L. Lasure, eds., p. 396-428: Academic Press: San Diego; and van den Hondel, C.A.M.J.J. & Punt, P.J. (1991) "Gene transfer systems and vector development for filamentous fungi, in: Applied Molecular Genetics of Fungi, Peberdy, J.F. et al., eds., p. 1-28, Cambridge University Press: Cambridge), algae and multicellular plant cells (see Schmidt, R. and Willmitzer, L. (1988) High efficiency *Agrobacterium tumefaciens* –mediated transformation of *Arabidopsis thaliana* leaf and cotyledon explants" *Plant Cell Rep.*: 583-586), or mammalian cells. Suitable host cells are discussed further in Goeddel, *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, CA (1990). Alternatively, the recombinant expression vector can be transcribed and translated *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase.

Expression of proteins in prokaryotes is most often carried out with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion proteins. Fusion vectors add a number of amino acids to a protein encoded therein, usually to the amino terminus of the recombinant protein but also to the C-terminus or fused within suitable regions in the proteins. Such fusion vectors typically serve three purposes: 1) to increase expression of recombinant protein; 2) to increase the solubility of the recombinant protein; and 3) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase.

Typical fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith, D.B. and Johnson, K.S. (1988) *Gene* 67:31-40), pMAL (New England Biolabs, Beverly, MA) and pRITS (Pharmacia, Piscataway, NJ) which fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein. In one embodiment, the coding sequence of the MCT protein is cloned into a pGEX expression vector to create a vector encoding a fusion protein comprising, from the N-terminus to the C-terminus, GST-thrombin cleavage site-X protein. The fusion protein can be purified by affinity chromatography using glutathione-agarose resin. Recombinant MCT protein unfused to GST can be recovered by cleavage of the fusion protein with thrombin.

Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amann et al., (1988) *Gene* 69:301-315) and pET 11d (Studier et al., *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, California (1990) 60-89). Target gene expression from the pTrc vector relies on host RNA polymerase transcription from a hybrid trp-lac fusion promoter. Target gene expression from the pET 11d vector relies on transcription from a T7 gn10-lac fusion promoter mediated by a coexpressed viral RNA polymerase (T7 gn1). This viral polymerase is supplied by host strains BL21(DE3) or HMS174(DE3) from a resident λ prophage harboring a T7 gn1 gene under the transcriptional control of the lacUV 5 promoter.

One strategy to maximize recombinant protein expression is to express the protein in a host bacteria with an impaired capacity to proteolytically cleave the recombinant protein (Gottesman, S., *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, California (1990) 119-128). Another strategy is to alter the nucleic acid sequence of the nucleic acid to be inserted into an expression vector so that the individual codons for each amino acid are those preferentially utilized in the bacterium chosen for expression, such as *C. glutamicum* (Wada et al. (1992) *Nucleic Acids Res.* 20:2111-2118). Such alteration of nucleic acid sequences of the invention can be carried out by standard DNA synthesis techniques.

In another embodiment, the MCT protein expression vector is a yeast expression vector. Examples of vectors for expression in yeast *S. cerivisae* include pYepSec1 (Baldari, et al., (1987) *Embo J.* 6:229-234), pMFa (Kurjan and Herskowitz, (1982) *Cell* 30:933-943), pJRY88 (Schultz et al., (1987) *Gene* 54:113-123), and pYES2 (Invitrogen Corporation, San Diego, CA). Vectors and methods for the construction of vectors appropriate for use in other fungi, such as the filamentous fungi, include those detailed in: van den Hondel, C.A.M.J.J. & Punt, P.J. (1991) "Gene transfer systems and vector

development for filamentous fungi, in: Applied Molecular Genetics of Fungi, J.F. Peberdy, et al., eds., p. 1-28, Cambridge University Press: Cambridge.

Alternatively, the MCT proteins of the invention can be expressed in insect cells using baculovirus expression vectors. Baculovirus vectors available for expression of proteins in cultured insect cells (e.g., Sf 9 cells) include the pAc series (Smith et al. (1983) *Mol. Cell Biol.* 3:2156-2165) and the pVL series (Lucklow and Summers (1989) *Virology* 170:31-39).

In another embodiment, the MCT proteins of the invention may be expressed in unicellular plant cells (such as algae) or in plant cells from higher plants (e.g., the spermatophytes, such as crop plants). Examples of plant expression vectors include those detailed in: Becker, D., Kemper, E., Schell, J. and Masterson, R. (1992) "New plant binary vectors with selectable markers located proximal to the left border", *Plant Mol. Biol.* 20: 1195-1197; and Bevan, M.W. (1984) "Binary *Agrobacterium* vectors for plant transformation", *Nucl Acid. Res.* 12: 8711-8721.

In yet another embodiment, a nucleic acid of the invention is expressed in mammalian cells using a mammalian expression vector. Examples of mammalian expression vectors include pCDM8 (Seed, B. (1987) *Nature* 329:840) and pMT2PC (Kaufman et al. (1987) *EMBO J.* 6:187-195). When used in mammalian cells, the expression vector's control functions are often provided by viral regulatory elements. For example, commonly used promoters are derived from polyoma, Adenovirus 2, cytomegalovirus and Simian Virus 40. For other suitable expression systems for both prokaryotic and eukaryotic cells see chapters 16 and 17 of Sambrook, J., Fritsh, E. F., and Maniatis, T. *Molecular Cloning: A Laboratory Manual. 2nd, ed., Cold Spring Harbor Laboratory*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.

In another embodiment, the recombinant mammalian expression vector is capable of directing expression of the nucleic acid preferentially in a particular cell type (e.g., tissue-specific regulatory elements are used to express the nucleic acid). Tissue-specific regulatory elements are known in the art. Non-limiting examples of suitable tissue-specific promoters include the albumin promoter (liver-specific; Pinkert et al. (1987) *Genes Dev.* 1:268-277), lymphoid-specific promoters (Calame and Eaton (1988) *Adv. Immunol* 43:235-275), in particular promoters of T cell receptors (Winoto and Baltimore (1989) *EMBO J.* 8:729-733) and immunoglobulins (Banerji et al. (1983) *Cell* 33:729-740; Queen and Baltimore (1983) *Cell* 33:741-748), neuron-specific promoters (e.g., the neurofilament promoter; Byrne and Ruddle (1989) *PNAS* 86:5473-5477), pancreas-specific promoters (Edlund et al. (1985) *Science* 230:912-916), and mammary gland-specific promoters (e.g., milk whey promoter; U.S. Patent No. 4,873,316 and

European Application Publication No. 264,166). Developmentally-regulated promoters are also encompassed, for example the murine hox promoters (Kessel and Gruss (1990) *Science* 249:374-379) and the α-fetoprotein promoter (Campes and Tilghman (1989) *Genes Dev.* 3:537-546).

The invention further provides a recombinant expression vector comprising a DNA molecule of the invention cloned into the expression vector in an antisense orientation. That is, the DNA molecule is operatively linked to a regulatory sequence in a manner which allows for expression (by transcription of the DNA molecule) of an RNA molecule which is antisense to MCT mRNA. Regulatory sequences operatively linked to a nucleic acid cloned in the antisense orientation can be chosen which direct the continuous expression of the antisense RNA molecule in a variety of cell types, for instance viral promoters and/or enhancers, or regulatory sequences can be chosen which direct constitutive, tissue specific or cell type specific expression of antisense RNA. The antisense expression vector can be in the form of a recombinant plasmid, phagemid or attenuated virus in which antisense nucleic acids are produced under the control of a high efficiency regulatory region, the activity of which can be determined by the cell type into which the vector is introduced. For a discussion of the regulation of gene expression using antisense genes see Weintraub, H. et al., Antisense RNA as a molecular tool for genetic analysis, *Reviews - Trends in Genetics*, Vol. 1(1) 1986.

Another aspect of the invention pertains to host cells into which a recombinant expression vector of the invention has been introduced. The terms "host cell" and "recombinant host cell" are used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

A host cell can be any prokaryotic or eukaryotic cell. For example, an MCT protein can be expressed in bacterial cells such as *C. glutamicum*, insect cells, yeast or mammalian cells (such as Chinese hamster ovary cells (CHO) or COS cells). Other suitable host cells are known to those skilled in the art. Microorganisms related to *Corynebacterium glutamicum* which may be conveniently used as host cells for the nucleic acid and protein molecules of the invention are set forth in Table 3.

Vector DNA can be introduced into prokaryotic or eukaryotic cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection", "conjugation" and "transduction" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid (e.g.,

DNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, natural competence, chemical-mediated transfer, or electroporation. Suitable methods for transforming or transfecting host cells can be found in Sambrook, et al. (*Molecular Cloning A Laboratory Manual. 2nd. ed., Cold Spring Harbor Laboratory*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989), and other laboratory manuals.

For stable transfection of mammalian cells, it is known that, depending upon the expression vector and transfection technique used, only a small fraction of cells may integrate the foreign DNA into their genome. In order to identify and select these integrants, a gene that encodes a selectable marker (e.g., resistance to antibiotics) is generally introduced into the host cells along with the gene of interest. Preferred selectable markers include those which confer resistance to drugs, such as G418, hygromycin and methotrexate. Nucleic acid encoding a selectable marker can be introduced into a host cell on the same vector as that encoding an MCT protein or can be introduced on a separate vector. Cells stably transfected with the introduced nucleic acid can be identified by, for example, drug selection (e.g., cells that have incorporated the selectable marker gene will survive, while the other cells die).

To create a homologous recombinant microorganism, a vector is prepared which contains at least a portion of an MCT gene into which a deletion, addition or substitution has been introduced to thereby alter, e.g., functionally disrupt, the MCT gene. Preferably, this MCT gene is a *Corynebacterium glutamicum* MCT gene, but it can be a homologue from a related bacterium or even from a mammalian, yeast, or insect source. In a preferred embodiment, the vector is designed such that, upon homologous recombination, the endogenous MCT gene is functionally disrupted (i.e., no longer encodes a functional protein; also referred to as a "knock out" vector). Alternatively, the vector can be designed such that, upon homologous recombination, the endogenous MCT gene is mutated or otherwise altered but still encodes functional protein (e.g., the upstream regulatory region can be altered to thereby alter the expression of the endogenous MCT protein). In the homologous recombination vector, the altered portion of the MCT gene is flanked at its 5' and 3' ends by additional nucleic acid of the MCT gene to allow for homologous recombination to occur between the exogenous MCT gene carried by the vector and an endogenous MCT gene in a microorganism. The additional flanking MCT nucleic acid is of sufficient length for successful homologous recombination with the endogenous gene. Typically, several kilobases of flanking DNA (both at the 5' and 3' ends) are included in the vector (see e.g., Thomas, K.R., and Capecchi, M.R. (1987) Cell 51: 503 for a description of homologous recombination vectors). The vector is introduced into a microorganism (e.g., by electroporation) and

cells in which the introduced MCT gene has homologously recombined with the endogenous MCT gene are selected, using art-known techniques.

In another embodiment, recombinant microorganisms can be produced which contain selected systems which allow for regulated expression of the introduced gene. For example, inclusion of an MCT gene on a vector placing it under control of the lac operon permits expression of the MCT gene only in the presence of IPTG. Such regulatory systems are well known in the art.

A host cell of the invention, such as a prokaryotic or eukaryotic host cell in culture, can be used to produce (i.e., express) an MCT protein. Accordingly, the invention further provides methods for producing MCT proteins using the host cells of the invention. In one embodiment, the method comprises culturing the host cell of invention (into which a recombinant expression vector encoding an MCT protein has been introduced, or into which genome has been introduced a gene encoding a wild-type or altered MCT protein) in a suitable medium until MCT protein is produced. In another embodiment, the method further comprises isolating MCT proteins from the medium or the host cell.

C. Isolated MCT Proteins

Another aspect of the invention pertains to isolated MCT proteins, and biologically active portions thereof. An "isolated" or "purified" protein or biologically active portion thereof is substantially free of cellular material when produced by recombinant DNA techniques, or chemical precursors or other chemicals when chemically synthesized. The language "substantially free of cellular material" includes preparations of MCT protein in which the protein is separated from cellular components of the cells in which it is naturally or recombinantly produced. In one embodiment, the language "substantially free of cellular material" includes preparations of MCT protein having less than about 30% (by dry weight) of non-MCT protein (also referred to herein as a "contaminating protein"), more preferably less than about 20% of non-MCT protein, still more preferably less than about 10% of non-MCT protein, and most preferably less than about 5% non-MCT protein. When the MCT protein or biologically active portion thereof is recombinantly produced, it is also preferably substantially free of culture medium, i.e., culture medium represents less than about 20%, more preferably less than about 10%, and most preferably less than about 5% of the volume of the protein preparation. The language "substantially free of chemical precursors or other chemicals" includes preparations of MCT protein in which the protein is separated from chemical precursors or other chemicals which are involved in the synthesis of the protein. In one embodiment, the language "substantially free of chemical precursors or

other chemicals" includes preparations of MCT protein having less than about 30% (by dry weight) of chemical precursors or non-MCT chemicals, more preferably less than about 20% chemical precursors or non-MCT chemicals, still more preferably less than about 10% chemical precursors or non-MCT chemicals, and most preferably less than about 5% chemical precursors or non-MCT chemicals. In preferred embodiments, isolated proteins or biologically active portions thereof lack contaminating proteins from the same organism from which the MCT protein is derived. Typically, such proteins are produced by recombinant expression of, for example, a *C. glutamicum* MCT protein in a microorganism such as *C. glutamicum*

An isolated MCT protein or a portion thereof of the invention can participate in the metabolism of compounds necessary for the construction of cellular membranes in *C. glutamicum*, or in the transport of molecules across these membranes, or has one or more of the activities set forth in Table 1. In preferred embodiments, the protein or portion thereof comprises an amino acid sequence which is sufficiently homologous to an amino acid sequence of Appendix B such that the protein or portion thereof maintains the ability participate in the metabolism of compounds necessary for the construction of cellular membranes in *C. glutamicum*, or in the transport of molecules across these membranes. The portion of the protein is preferably a biologically active portion as described herein. In another preferred embodiment, an MCT protein of the invention has an amino acid sequence shown in Appendix B. In yet another preferred embodiment, the MCT protein has an amino acid sequence which is encoded by a nucleotide sequence which hybridizes, e.g., hybridizes under stringent conditions, to a nucleotide sequence of Appendix A. In still another preferred embodiment, the MCT protein has an amino acid sequence which is encoded by a nucleotide sequence that is at least about 50-60%, preferably at least about 60-70%, more preferably at least about 70-80%, 80-90%, 90-95%, and even more preferably at least about 96%, 97%, 98%, 99% or more homologous to one of the amino acid sequences of Appendix B. The preferred MCT proteins of the present invention also preferably possess at least one of the MCT activities described herein. For example, a preferred MCT protein of the present invention includes an amino acid sequence encoded by a nucleotide sequence which hybridizes, e.g., hybridizes under stringent conditions, to a nucleotide sequence of Appendix A, and which can participate in the metabolism of compounds necessary for the construction of cellular membranes in *C. glutamicum*, or in the transport of molecules across these membranes, or which has one or more of the activities set forth in Table 1.

In other embodiments, the MCT protein is substantially homologous to an amino acid sequence of Appendix B and retains the functional activity of the protein of one of

the sequences of Appendix B yet differs in amino acid sequence due to natural variation or mutagenesis, as described in detail in subsection I above. Accordingly, in another embodiment, the MCT protein is a protein which comprises an amino acid sequence which is at least about 50-60%, preferably at least about 60-70%, and more preferably at least about 70-80, 80-90, 90-95%, and most preferably at least about 96%, 97%, 98%, 99% or more homologous to an entire amino acid sequence of Appendix B and which has at least one of the MCT activities described herein. In another embodiment, the invention pertains to a full length *C. glutamicum* protein which is substantially homologous to an entire amino acid sequence of Appendix B.

Biologically active portions of an MCT protein include peptides comprising amino acid sequences derived from the amino acid sequence of an MCT protein, e.g., the an amino acid sequence shown in Appendix B or the amino acid sequence of a protein homologous to an MCT protein, which include fewer amino acids than a full length MCT protein or the full length protein which is homologous to an MCT protein, and exhibit at least one activity of an MCT protein. Typically, biologically active portions (peptides, e.g., peptides which are, for example, 5, 10, 15, 20, 30, 35, 36, 37, 38, 39, 40, 50, 100 or more amino acids in length) comprise a domain or motif with at least one activity of an MCT protein. Moreover, other biologically active portions, in which other regions of the protein are deleted, can be prepared by recombinant techniques and evaluated for one or more of the activities described herein. Preferably, the biologically active portions of an MCT protein include one or more selected domains/motifs or portions thereof having biological activity.

MCT proteins are preferably produced by recombinant DNA techniques. For example, a nucleic acid molecule encoding the protein is cloned into an expression vector (as described above), the expression vector is introduced into a host cell (as described above) and the MCT protein is expressed in the host cell. The MCT protein can then be isolated from the cells by an appropriate purification scheme using standard protein purification techniques. Alternative to recombinant expression, an MCT protein, polypeptide, or peptide can be synthesized chemically using standard peptide synthesis techniques. Moreover, native MCT protein can be isolated from cells (e.g., endothelial cells), for example using an anti-MCT antibody, which can be produced by standard techniques utilizing an MCT protein or fragment thereof of this invention.

The invention also provides MCT chimeric or fusion proteins. As used herein, an MCT "chimeric protein" or "fusion protein" comprises an MCT polypeptide operatively linked to a non-MCT polypeptide. An "MCT polypeptide" refers to a polypeptide having an amino acid sequence corresponding to an MCT protein, whereas a "non-MCT polypeptide" refers to a polypeptide having an amino acid sequence

corresponding to a protein which is not substantially homologous to the MCT protein, e.g., a protein which is different from the MCT protein and which is derived from the same or a different organism. Within the fusion protein, the term "operatively linked" is intended to indicate that the MCT polypeptide and the non-MCT polypeptide are fused in-frame to each other. The non-MCT polypeptide can be fused to the N-terminus or C-terminus of the MCT polypeptide. For example, in one embodiment the fusion protein is a GST-MCT fusion protein in which the MCT sequences are fused to the C-terminus of the GST sequences. Such fusion proteins can facilitate the purification of recombinant MCT proteins. In another embodiment, the fusion protein is an MCT protein containing a heterologous signal sequence at its N-terminus. In certain host cells (e.g., mammalian host cells), expression and/or secretion of an MCT protein can be increased through use of a heterologous signal sequence.

Preferably, an MCT chimeric or fusion protein of the invention is produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different polypeptide sequences are ligated together in-frame in accordance with conventional techniques, for example by employing blunt-ended or stagger-ended termini for ligation, restriction enzyme digestion to provide for appropriate termini, filling-in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid undesirable joining, and enzymatic ligation. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers which give rise to complementary overhangs between two consecutive gene fragments which can subsequently be annealed and reamplified to generate a chimeric gene sequence (see, for example, *Current Protocols in Molecular Biology*, eds. Ausubel et al. John Wiley & Sons: 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (e.g., a GST polypeptide). An MCT-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the MCT protein.

Homologues of the MCT protein can be generated by mutagenesis, e.g., discrete point mutation or truncation of the MCT protein. As used herein, the term "homologue" refers to a variant form of the MCT protein which acts as an agonist or antagonist of the activity of the MCT protein. An agonist of the MCT protein can retain substantially the same, or a subset, of the biological activities of the MCT protein. An antagonist of the MCT protein can inhibit one or more of the activities of the naturally occurring form of the MCT protein, by, for example, competitively binding to a downstream or upstream member of the cell membrane component metabolic cascade which includes the MCT

protein, or by binding to an MCT protein which mediates transport of compounds across such membranes, thereby preventing translocation from taking place.

In an alternative embodiment, homologues of the MCT protein can be identified by screening combinatorial libraries of mutants, e.g., truncation mutants, of the MCT protein for MCT protein agonist or antagonist activity. In one embodiment, a variegated library of MCT variants is generated by combinatorial mutagenesis at the nucleic acid level and is encoded by a variegated gene library. A variegated library of MCT variants can be produced by, for example, enzymatically ligating a mixture of synthetic oligonucleotides into gene sequences such that a degenerate set of potential MCT sequences is expressible as individual polypeptides, or alternatively, as a set of larger fusion proteins (e.g., for phage display) containing the set of MCT sequences therein. There are a variety of methods which can be used to produce libraries of potential MCT homologues from a degenerate oligonucleotide sequence. Chemical synthesis of a degenerate gene sequence can be performed in an automatic DNA synthesizer, and the synthetic gene then ligated into an appropriate expression vector. Use of a degenerate set of genes allows for the provision, in one mixture, of all of the sequences encoding the desired set of potential MCT sequences. Methods for synthesizing degenerate oligonucleotides are known in the art (see, e.g., Narang, S.A. (1983) *Tetrahedron* 39:3; Itakura et al. (1984) *Annu. Rev. Biochem.* 53:323; Itakura et al. (1984) *Science* 198:1056; Ike et al. (1983) *Nucleic Acid Res.* 11:477.

In addition, libraries of fragments of the MCT protein coding can be used to generate a variegated population of MCT fragments for screening and subsequent selection of homologues of an MCT protein. In one embodiment, a library of coding sequence fragments can be generated by treating a double stranded PCR fragment of an MCT coding sequence with a nuclease under conditions wherein nicking occurs only about once per molecule, denaturing the double stranded DNA, renaturing the DNA to form double stranded DNA which can include sense/antisense pairs from different nicked products, removing single stranded portions from reformed duplexes by treatment with S1 nuclease, and ligating the resulting fragment library into an expression vector. By this method, an expression library can be derived which encodes N-terminal, C-terminal and internal fragments of various sizes of the MCT protein.

Several techniques are known in the art for screening gene products of combinatorial libraries made by point mutations or truncation, and for screening cDNA libraries for gene products having a selected property. Such techniques are adaptable for rapid screening of the gene libraries generated by the combinatorial mutagenesis of MCT homologues. The most widely used techniques, which are amenable to high through-put analysis, for screening large gene libraries typically include cloning the

gene library into replicable expression vectors, transforming appropriate cells with the resulting library of vectors, and expressing the combinatorial genes under conditions in which detection of a desired activity facilitates isolation of the vector encoding the gene whose product was detected. Recursive ensemble mutagenesis (REM), a new technique which enhances the frequency of functional mutants in the libraries, can be used in combination with the screening assays to identify MCT homologues (Arkin and Yourvan (1992) *PNAS* 89:7811-7815; Delgrave et al. (1993) *Protein Engineering* 6(3):327-331).

In another embodiment, cell based assays can be exploited to analyze a variegated MCT library, using methods well known in the art.

D. Uses and Methods of the Invention

The nucleic acid molecules, proteins, protein homologues, fusion proteins, primers, vectors, and host cells described herein can be used in one or more of the following methods: identification of *C. glutamicum* and related organisms; mapping of genomes of organisms related to *C. glutamicum*; identification and localization of *C. glutamicum* sequences of interest; evolutionary studies; determination of MCT protein regions required for function; modulation of an MCT protein activity; modulation of the metabolism of one or more cell membrane components; modulation of the transmembrane transport of one or more compounds; and modulation of cellular production of a desired compound, such as a fine chemical.

The MCT nucleic acid molecules of the invention have a variety of uses. First, they may be used to identify an organism as being *Corynebacterium glutamicum* or a close relative thereof. Also, they may be used to identify the presence of *C glutamicum* or a relative thereof in a mixed population of microorganisms. The invention provides the nucleic acid sequences of a number of *C. glutamicum* genes; by probing the extracted genomic DNA of a culture of a unique or mixed population of microorganisms under stringent conditions with a probe spanning a region of a *C. glutamicum* gene which is unique to this organism, one can ascertain whether this organism is present. Although *Corynebacterium glutamicum* itself is nonpathogenic, it is related to pathogenic species, such as *Corynebacterium diphtheriae*. Detection of such organisms is of significant clinical relevance.

Further, the nucleic acid and protein molecules of the invention may serve as markers for specific regions of the genome. This has utility not only in the mapping of the genome, but also for functional studies of *C. glutamicum* proteins. For example, to identify the region of the genome to which a particular *C. glutamicum* DNA-binding protein binds, the *C. glutamicum* genome could be digested, and the fragments incubated

with the DNA-binding protein. Those which bind the protein may be additionally probed with the nucleic acid molecules of the invention, preferably with readily detectable labels; binding of such a nucleic acid molecule to the genome fragment enables the localization of the fragment to the genome map of *C. glutamicum*, and, when performed multiple times with different enzymes, facilitates a rapid determination of the nucleic acid sequence to which the protein binds. Further, the nucleic acid molecules of the invention may be sufficiently homologous to the sequences of related species such that these nucleic acid molecules may serve as markers for the construction of a genomic map in related bacteria, such as *Brevibacterium lactofermentum*.

The MCT nucleic acid molecules of the invention are also useful for evolutionary and protein structural studies. The metabolic and transport processes in which the molecules of the invention participate are utilized by a wide variety of prokaryotic and eukaryotic cells; by comparing the sequences of the nucleic acid molecules of the present invention to those encoding similar enzymes from other organisms, the evolutionary relatedness of the organisms can be assessed. Similarly, such a comparison permits an assessment of which regions of the sequence are conserved and which are not, which may aid in determining those regions of the protein which are essential for the functioning of the enzyme. This type of determination is of value for protein engineering studies and may give an indication of what the protein can tolerate in terms of mutagenesis without losing function.

Manipulation of the MCT nucleic acid molecules of the invention may result in the production of MCT proteins having functional differences from the wild-type MCT proteins. These proteins may be improved in efficiency or activity, may be present in greater numbers in the cell than is usual, or may be decreased in efficiency or activity.

There are a number of mechanisms by which the alteration of an MCT protein of the invention may directly affect the yield, production, and/or efficiency of production of a fine chemical from a *C. glutamicum* strain incorporating such an altered protein. Recovery of fine chemical compounds from large-scale cultures of *C. glutamicum* is significantly improved if *C. glutamicum* secretes the desired compounds, since such compounds may be readily purified from the culture medium (as opposed to extracted from the mass of *C glutamicum* cells). By either increasing the number or the activity of transporter molecules which export fine chemicals from the cell, it may be possible to increase the amount of the produced fine chemical which is present in the extracellular medium, thus permitting greater ease of harvesting and purification. Conversely, in order to efficiently overproduce one or more fine chemicals, increased amounts of the cofactors, precursor molecules, and intermediate compounds for the appropriate biosynthetic pathways are required. Therefore, by increasing the number and/or activity

of transporter proteins involved in the import of nutrients, such as carbon sources (i.e., sugars), nitrogen sources (i.e., amino acids, ammonium salts), phosphate, and sulfur, it may be possible to improve the production of a fine chemical, due to the removal of any nutrient supply limitations on the biosynthetic process. Further, fatty acids and lipids are themselves desirable fine chemicals, so by optimizing the activity or increasing the number of one or more MCT proteins of the invention which participate in the biosynthesis of these compounds, or by impairing the activity of one or more MCT proteins which are involved in the degradation of these compounds, it may be possible to increase the yield, production, and/or efficiency of production of fatty acid and lipid molecules from *C. glutamicum*.

The engineering of one or more MCT genes of the invention may also result in MCT proteins having altered activities which indirectly impact the production of one or more desired fine chemicals from *C .glutamicum*. For example, the normal biochemical processes of metabolism result in the production of a variety of waste products (e.g., hydrogen peroxide and other reactive oxygen species) which may actively interfere with these same metabolic processes (for example, peroxynitrite is known to nitrate tyrosine side chains, thereby inactivating some enzymes having tyrosine in the active site (Groves, J.T. (1999) *Curr. Opin. Chem. Biol.* 3(2): 226-235). While these waste products are typically excreted, the *C. glutamicum* strains utilized for large-scale fermentative production are optimized for the overproduction of one or more fine chemicals, and thus may produce more waste products than is typical for a wild-type *C. glutamicum*. By optimizing the activity of one or more MCT proteins of the invention which are involved in the export of waste molecules, it may be possible to improve the viability of the cell and to maintain efficient metabolic activity. Also, the presence of high intracellular levels of the desired fine chemical may actually be toxic to the cell, so by increasing the ability of the cell to secrete these compounds, one may improve the viability of the cell.

Further, the MCT proteins of the invention may be manipulated such that the relative amounts of various lipid and fatty acid molecules produced are altered. This may have a profound effect on the lipid composition of the membrane of the cell. Since each type of lipid has different physical properties, an alteration in the lipid composition of a membrane may significantly alter membrane fluidity. Changes in membrane fluidity can impact the transport of molecules across the membrane, which, as previously explicated, may modify the export of waste products or the produced fine chemical or the import of necessary nutrients. Such membrane fluidity changes may also profoundly affect the integrity of the cell; cells with relatively weaker membranes are more vulnerable in the large-scale fermentor environment to mechanical stresses which may

damage or kill the cell. By manipulating MCT proteins involved in the production of fatty acids and lipids for membrane construction such that the resulting membrane has a membrane composition more amenable to the environmental conditions extant in the cultures utilized to produce fine chemicals, a greater proportion of the *C. glutamicum* cells should survive and multiply. Greater numbers of *C. glutamicum* cells in a culture should translate into greater yields, production, or efficiency of production of the fine chemical from the culture.

The aforementioned mutagenesis strategies for MCT proteins to result in increased yields of a fine chemical from *C. glutamicum* are not meant to be limiting; variations on these strategies will be readily apparent to one skilled in the art. Using such strategies, and incorporating the mechanisms disclosed herein, the nucleic acid and protein molecules of the invention may be utilized to generate *C. glutamicum* or related strains of bacteria expressing mutated MCT nucleic acid and protein molecules such that the yield, production, and/or efficiency of production of a desired compound is improved. This desired compound may be any natural product of *C. glutamicum*, which includes the final products of biosynthesis pathways and intermediates of naturally-occurring metabolic pathways, as well as molecules which do not naturally occur in the metabolism of *C. glutamicum*, but which are produced by a *C. glutamicum* strain of the invention.

This invention is further illustrated by the following examples which should not be construed as limiting. The contents of all references, patent applications, patents, and published patent applications cited throughout this application are hereby incorporated by reference.

Exemplification

Example 1: Preparation of total genomic DNA of *Corynebacterium glutamicum* ATCC 13032

A culture of *Corynebacterium glutamicum* (ATCC 13032) was grown overnight at 30°C with vigorous shaking in BHI medium (Difco). The cells were harvested by centrifugation, the supernatant was discarded and the cells were resuspended in 5 ml buffer-I (5% of the original volume of the culture — all indicated volumes have been calculated for 100 ml of culture volume). Composition of buffer-I: 140.34 g/l sucrose, 2.46 g/l $MgSO_4 \times 7H_2O$, 10 ml/l KH_2PO_4 solution (100 g/l, adjusted to pH 6.7 with KOH), 50 ml/l M12 concentrate (10 g/l $(NH_4)_2SO_4$, 1 g/l NaCl, 2 g/l $MgSO_4 \times 7H_2O$, 0.2 g/l $CaCl_2$, 0.5 g/l yeast extract (Difco), 10 ml/l trace-elements-mix (200 mg/l $FeSO_4 \times H_2O$, 10 mg/l $ZnSO_4 \times 7\ H_2O$, 3 mg/l $MnCl_2 \times 4\ H_2O$, 30 mg/l H_3BO_3, 20 mg/l $CoCl_2 \times 6\ H_2O$, 1 mg/l $NiCl_2 \times 6\ H_2O$, 3 mg/l $Na_2MoO_4 \times 2\ H_2O$, 500 mg/l complexing agent (EDTA or critic acid), 100 ml/l vitamins-mix (0.2 mg/l biotin, 0.2 mg/l folic acid, 20 mg/l p-amino benzoic acid, 20 mg/l riboflavin, 40 mg/l ca-panthothenate, 140 mg/l nicotinic acid, 40 mg/l pyridoxole hydrochloride, 200 mg/l myo-inositol). Lysozyme was added to the suspension to a final concentration of 2.5 mg/ml. After an approximately 4 h incubation at 37°C, the cell wall was degraded and the resulting protoplasts are harvested by centrifugation. The pellet was washed once with 5 ml buffer-I and once with 5 ml TE-buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8). The pellet was resuspended in 4 ml TE-buffer and 0.5 ml SDS solution (10%) and 0.5 ml NaCl solution (5 M) are added. After adding of proteinase K to a final concentration of 200 µg/ml, the suspension is incubated for ca.18 h at 37°C. The DNA was purified by extraction with phenol, phenol-chloroform-isoamylalcohol and chloroform-isoamylalcohol using standard procedures. Then, the DNA was precipitated by adding 1/50 volume of 3 M sodium acetate and 2 volumes of ethanol, followed by a 30 min incubation at -20°C and a 30 min centrifugation at 12,000 rpm in a high speed centrifuge using a SS34 rotor (Sorvall). The DNA was dissolved in 1 ml TE-buffer containing 20

µg/ml RNaseA and dialysed at 4°C against 1000 ml TE-buffer for at least 3 hours. During this time, the buffer was exchanged 3 times. To aliquots of 0.4 ml of the dialysed DNA solution, 0.4 ml of 2 M LiCl and 0.8 ml of ethanol are added. After a 30 min incubation at -20°C, the DNA was collected by centrifugation (13,000 rpm, Biofuge Fresco, Heraeus, Hanau, Germany). The DNA pellet was dissolved in TE-buffer. DNA prepared by this procedure could be used for all purposes, including southern blotting or construction of genomic libraries.

Example 2: Construction of genomic libraries in *Escherichia coli* of *Corynebacterium glutamicum* ATCC13032.

Starting from DNA prepared as described in Example 1, cosmid and plasmid libraries were constructed according to known and well established methods (*see* e.g., Sambrook, J. *et al.* (1989) "Molecular Cloning : A Laboratory Manual", Cold Spring Harbor Laboratory Press, or Ausubel, F.M. *et al.* (1994) "Current Protocols in Molecular Biology", John Wiley & Sons.)

Any plasmid or cosmid could be used. Of particular use were the plasmids pBR322 (Sutcliffe, J.G. (1979) Proc. Natl. Acad. Sci. USA, 75:3737-3741); pACYC177 (Change & Cohen (1978) J. Bacteriol 134:1141-1156), plasmids of the pBS series (pBSSK+, pBSSK- and others; Stratagene, LaJolla, USA), or cosmids as SuperCos1 (Stratagene, LaJolla, USA) or Lorist6 (Gibson, T.J., Rosenthal A. and Waterson, R.H. (1987) Gene 53:283-286.

Example 3: DNA Sequencing and Computational Functional Analysis

Genomic libraries as described in Example 2 were used for DNA sequencing according to standard methods, in particular by the chain termination method using ABI377 sequencing machines (see e.g., Fleischman, R.D. et al. (1995) "Whole-genome Random Sequencing and Assembly of Haemophilus Influenzae Rd., Science, 269:496-512). Sequencing primers with the following nucleotide sequences were used: 5'-GGAAACAGTATGACCATG-3' or 5'-GTAAAACGACGGCCAGT-3'.

Example 4: *In vivo* Mutagenesis

In vivo mutagenesis of *Corynebacterium glutamicum* can be performed by passage of plasmid (or other vector) DNA through *E. coli* or other microorganisms (e.g. *Bacillus* spp. or yeasts such as *Saccharomyces cerevisiae*) which are impaired in their capabilities to maintain

the integrity of their genetic information. Typical mutator strains have mutations in the genes for the DNA repair system (e.g., mutHLS, mutD, mutT, etc.; for reference, see Rupp, W.D. (1996) DNA repair mechanisms, in: *Escherichia coli* and *Salmonella*, p. 2277-2294, ASM: Washington.) Such strains are well known to those skilled in the art. The use of such strains is illustrated, for example, in Greener, A. and Callahan, M. (1994) Strategies 7: 32-34.

Example 5: DNA Transfer Between *Escherichia coli* and *Corynebacterium glutamicum*

Several *Corynebacterium* and *Brevibacterium* species contain endogenous plasmids (as e.g., pHM1519 or pBL1) which replicate autonomously (for review see, e.g., Martin, J.F. *et al* (1987) Biotechnology, 5:137-146). Shuttle vectors for *Escherichia coli* and *Corynebacterium glutamicum* can be readily constructed by using standard vectors for *E coli* (Sambrook, J. *et al.* (1989), "Molecular Cloning: A Laboratory Manual", Cold Spring Harbor Laboratory Press or Ausubel, F.M. *et al.* (1994) "Current Protocols in Molecular Biology", John Wiley & Sons) to which a origin or replication for and a suitable marker from *Corynebacterium glutamicum* is added. Such origins of replication are preferably taken from endogenous plasmids isolated from *Corynebacterium* and *Brevibacterium* species. Of particular use as transformation markers for these species are genes for kanamycin resistance (such as those derived from the Tn5 or Tn903 transposons) or chloramphenicol (Winnacker, E.L. (1987) "From Genes to Clones — Introduction to Gene Technology, VCH, Weinheim). There are numerous examples in the literature of the construction of a wide variety of shuttle vectors which replicate in both *E. coli and C. glutamicum*, and which can be used for several purposes, including gene over-expression (for reference, see e.g., Yoshihama, M. et al. (1985) J. Bacteriol. 162:591-597, Martin J.F. et al. (1987) Biotechnology, 5:137-146 and Eikmanns, B.J. et al. (1991) Gene, 102:93-98).

Using standard methods, it is possible to clone a gene of interest into one of the shuttle vectors described above and to introduce such a hybrid vectors into strains of *Corynebacterium glutamicum*. Transformation of *C glutamicum* can be achieved by protoplast transformation (Kastsumata, R. et al. (1984) J. Bacteriol. 159306-311), electroporation (Liebl, E. et al. (1989) FEMS Microbiol. Letters, 53:399-303) and in cases where special vectors are used, also by conjugation (as described e.g. in Schäfer, A et al. (1990) J. Bacteriol. 172:1663-1666). It is also possible to transfer the shuttle vectors for

C. glutamicum to *E. coli* by preparing plasmid DNA from *C. glutamicum* (using standard methods well-known in the art) and transforming it into *E. coli*. This transformation step can be performed using standard methods, but it is advantageous to use an Mcr-deficient *E coli* strain, such as NM522 (Gough & Murray (1983) J. Mol. Biol. 166:1-19).

Example 6: Assessment of the Expression of the Mutant Protein

Observations of the activity of a mutated protein in a transformed host cell rely on the fact that the mutant protein is expressed in a similar fashion and in a similar quantity to that of the wild-type protein. A useful method to ascertain the level of transcription of the mutant gene (an indicator of the amount of mRNA available for translation to the gene product) is to perform a Northern blot (for reference see, for example, Ausubel et al. (1988) Current Protocols in Molecular Biology, Wiley: New York), in which a primer designed to bind to the gene of interest is labeled with a detectable tag (usually radioactive or chemiluminescent), such that when the total RNA of a culture of the organism is extracted, run on gel, transferred to a stable matrix and incubated with this probe, the binding and quantity of binding of the probe indicates the presence and also the quantity of mRNA for this gene. This information is evidence of the degree of transcription of the mutant gene. Total cellular RNA can be prepared from *Corynebacterium glutamicum* by several methods, all well-known in the art, such as that described in Bormann, E.R. et al. (1992) *Mol. Microbiol.* 6: 317-326.

To assess the presence or relative quantity of protein translated from this mRNA, standard techniques, such as a Western blot, may be employed (see, for example, Ausubel et al. (1988) Current Protocols in Molecular Biology, Wiley: New York). In this process, total cellular proteins are extracted, separated by gel electrophoresis, transferred to a matrix such as nitrocellulose, and incubated with a probe, such as an antibody, which specifically binds to the desired protein. This probe is generally tagged with a chemiluminescent or colorimetric label which may be readily detected. The presence and quantity of label observed indicates the presence and quantity of the desired mutant protein present in the cell.

Example 7: Growth of Genetically Modified *Corynebacterium glutamicum* — Media and Culture Conditions

Genetically modified *Corynebacteria* are cultured in synthetic or natural growth media. A number of different growth media for Corynebacteria are both well-known and readily available (Lieb *et al.* (1989) Appl. Microbiol. Biotechnol., 32:205-210; von der Osten *et al.* (1998) Biotechnology Letters, 11:11-16; Patent DE 4,120,867; Liebl (1992) "The Genus *Corynebacterium*, in: The Procaryotes, Volume II, Balows, A. *et al.*, eds. Springer-Verlag). These media consist of one or more carbon sources, nitrogen sources, inorganic salts, vitamins and trace elements. Preferred carbon sources are sugars, such as mono-, di-, or polysaccharides. For example, glucose, fructose, mannose, galactose, ribose, sorbose, ribulose, lactose, maltose, sucrose, raffinose, starch or cellulose serve as very good carbon sources. It is also possible to supply sugar to the media via complex compounds such as molasses or other by-products from sugar refinement. It can also be advantageous to supply mixtures of different carbon sources. Other possible carbon sources are alcohols and organic acids, such as methanol, ethanol, acetic acid or lactic acid. Nitrogen sources are usually organic or inorganic nitrogen compounds, or materials which contain these compounds. Exemplary nitrogen sources include ammonia gas or ammonia salts, such as NH_4Cl or $(NH_4)_2SO_4$, NH_4OH, nitrates, urea, amino acids or complex nitrogen sources like corn steep liquor, soy bean flour, soy bean protein, yeast extract, meat extract and others.

Inorganic salt compounds which may be included in the media include the chloride-, phosphorous- or sulfate- salts of calcium, magnesium, sodium, cobalt, molybdenum, potassium, manganese, zinc, copper and iron. Chelating compounds can be added to the medium to keep the metal ions in solution. Particularly useful chelating compounds include dihydroxyphenols, like catechol or protocatechuate, or organic acids, such as citric acid. It is typical for the media to also contain other growth factors, such as vitamins or growth promoters, examples of which include biotin, riboflavin, thiamin, folic acid, nicotinic acid, pantothenate and pyridoxin. Growth factors and salts frequently originate from complex media components such as yeast extract, molasses, corn steep liquor and others. The exact composition of the media compounds depends strongly on the immediate experiment and is individually decided for each specific case. Information about media optimization is available in the textbook "Applied Microbiol. Physiology, A Practical Approach (*eds.* P.M. Rhodes, P.F. Stanbury, IRL Press (1997) pp. 53-73, ISBN 0

19 963577 3). It is also possible to select growth media from commercial suppliers, like standard 1 (Merck) or BHI (grain heart infusion, DIFC) or others.

All medium components are sterilized, either by heat (20 minutes at 1.5 bar and 121°C) or by sterile filtration. The components can either be sterilized together or, if necessary, separately. All media components can be present at the beginning of growth, or they can optionally be added continuously or batchwise.

Culture conditions are defined separately for each experiment. The temperature should be in a range between 15°C and 45°C. The temperature can be kept constant or can be altered during the experiment. The pH of the medium should be in the range of 5 to 8.5, preferably around 7.0, and can be maintained by the addition of buffers to the media. An exemplary buffer for this purpose is a potassium phosphate buffer. Synthetic buffers such as MOPS, HEPES, ACES and others can alternatively or simultaneously be used. It is also possible to maintain a constant culture pH through the addition of NaOH or NH_4OH during growth. If complex medium components such as yeast extract are utilized, the necessity for additional buffers may be reduced, due to the fact that many complex compounds have high buffer capacities. If a fermentor is utilized for culturing the micro-organisms, the pH can also be controlled using gaseous ammonia.

The incubation time is usually in a range from several hours to several days. This time is selected in order to permit the maximal amount of product to accumulate in the broth. The disclosed growth experiments can be carried out in a variety of vessels, such as microtiter plates, glass tubes, glass flasks or glass or metal fermentors of different sizes. For screening a large number of clones, the microorganisms should be cultured in microtiter plates, glass tubes or shake flasks, either with or without baffles. Preferably 100 ml shake flasks are used, filled with 10% (by volume) of the required growth medium. The flasks should be shaken on a rotary shaker (amplitude 25 mm) using a speed-range of 100 – 300 rpm. Evaporation losses can be diminished by the maintenance of a humid atmosphere; alternatively, a mathematical correction for evaporation losses should be performed.

If genetically modified clones are tested, an unmodified control clone or a control clone containing the basic plasmid without any insert should also be tested. The medium is inoculated to an OD_{600} of 0.5 – 1.5 using cells grown on agar plates, such as CM plates (10 g/l glucose, 2,5 g/l NaCl, 2 g/l urea, 10 g/l polypeptone, 5 g/l yeast extract, 5 g/l meat extract, 22 g/l NaCl, 2 g/l urea, 10 g/l polypeptone, 5 g/l yeast extract, 5 g/l meat extract,

22 g/l agar, pH 6.8 with 2M NaOH) that had been incubated at 30°C. Inoculation of the media is accomplished by either introduction of a saline suspension of *C. glutamicum* cells from CM plates or addition of a liquid preculture of this bacterium.

Example 8 – *In vitro* Analysis of the Function of Mutant Proteins

The determination of activities and kinetic parameters of enzymes is well established in the art. Experiments to determine the activity of any given altered enzyme must be tailored to the specific activity of the wild-type enzyme, which is well within the ability of one skilled in the art. Overviews about enzymes in general, as well as specific details concerning structure, kinetics, principles, methods, applications and examples for the determination of many enzyme activities may be found, for example, in the following references: Dixon, M., and Webb, E.C., (1979) Enzymes. Longmans: London; Fersht, (1985) Enzyme Structure and Mechanism. Freeman: New York; Walsh, (1979) Enzymatic Reaction Mechanisms. Freeman: San Francisco; Price, N.C., Stevens, L. (1982) Fundamentals of Enzymology. Oxford Univ. Press: Oxford; Boyer, P.D., ed. (1983) The Enzymes, 3rd ed. Academic Press: New York; Bisswanger, H., (1994) Enzymkinetik, 2nd ed. VCH: Weinheim (ISBN 3527300325); Bergmeyer, H.U., Bergmeyer, J., Graßl, M., eds. (1983-1986) Methods of Enzymatic Analysis, 3rd ed., vol. I-XII, Verlag Chemie: Weinheim; and Ullmann's Encyclopedia of Industrial Chemistry (1987) vol. A9, "Enzymes". VCH: Weinheim, p. 352-363.

The activity of proteins which bind to DNA can be measured by several well-established methods, such as DNA band-shift assays (also called gel retardation assays). The effect of such proteins on the expression of other molecules can be measured using reporter gene assays (such as that described in Kolmar, H. et al. (1995) EMBO J. 14: 3895-3904 and references cited therein). Reporter gene test systems are well known and established for applications in both pro- and eukaryotic cells, using enzymes such as beta-galactosidase, green fluorescent protein, and several others.

The determination of activity of membrane-transport proteins can be performed according to techniques such as those described in Gennis, R.B. (1989) "Pores, Channels and Transporters", in Biomembranes, Molecular Structure and Function, Springer: Heidelberg, p. 85-137; 199-234; and 270-322.

Example 9: Analysis of Impact of Mutant Protein on the Production of the Desired Product

The effect of the genetic modification in *C. glutamicum* on production of a desired compound (such as an amino acid) can be assessed by growing the modified microorganism under suitable conditions (such as those described above) and analyzing

the medium and/or the cellular component for increased production of the desired product (i.e., an amino acid). Such analysis techniques are well known to one skilled in the art, and include spectroscopy, thin layer chromatography, staining methods of various kinds, enzymatic and microbiological methods, and analytical chromatography such as high performance liquid chromatography (see, for example, Ullman, Encyclopedia of Industrial Chemistry, vol. A2, p. 89-90 and p. 443-613, VCH: Weinheim (1985); Fallon, A. et al., (1987) "Applications of HPLC in Biochemistry" in: Laboratory Techniques in Biochemistry and Molecular Biology, vol. 17; Rehm et al. (1993) Biotechnology, vol. 3, Chapter III: "Product recovery and purification", page 469-714, VCH: Weinheim; Belter, P.A. et al. (1988) Bioseparations: downstream processing for biotechnology, John Wiley and Sons; Kennedy, J.F. and Cabral, J.M.S. (1992) Recovery processes for biological materials, John Wiley and Sons; Shaeiwitz, J.A. and Henry, J.D. (1988) Biochemical separations, in: Ulmann's Encyclopedia of Industrial Chemistry, vol. B3, Chapter 11, page 1-27, VCH: Weinheim; and Dechow, F.J. (1989) Separation and purification techniques in biotechnology, Noyes Publications.)

In addition to the measurement of the final product of fermentation, it is also possible to analyze other components of the metabolic pathways utilized for the production of the desired compound, such as intermediates and side-products, to determine the overall efficiency of production of the compound. Analysis methods include measurements of nutrient levels in the medium (e.g., sugars, hydrocarbons, nitrogen sources, phosphate, and other ions), measurements of biomass composition and growth, analysis of the production of common metabolites of biosynthetic pathways, and measurement of gasses produced during fermentation. Standard methods for these measurements are outlined in Applied Microbial Physiology, A Practical Approach, P.M. Rhodes and P.F. Stanbury, eds., IRL Press, p. 103-129; 131-163; and 165-192 (ISBN: 0199635773) and references cited therein.

Example 10: Purification of the Desired Product from *C. glutamicum* Culture

Recovery of the desired product from the *C. glutamicum* cells or supernatant of the above-described culture can be performed by various methods well known in the art. If the desired product is not secreted from the cells, the cells can be harvested from the culture by low-speed centrifugation, the cells can be lysed by standard techniques, such as mechanical force or sonication. The cellular debris is removed by centrifugation, and the supernatant fraction containing the soluble proteins is retained for further purification of the desired compound. If the product is secreted from the *C. glutamicum*

cells, then the cells are removed from the culture by low-speed centrifugation, and the supernate fraction is retained for further purification.

The supernatant fraction from either purification method is subjected to chromatography with a suitable resin, in which the desired molecule is either retained on a chromatography resin while many of the impurities in the sample are not, or where the impurities are retained by the resin while the sample is not. Such chromatography steps may be repeated as necessary, using the same or different chromatography resins. One skilled in the art would be well-versed in the selection of appropriate chromatography resins and in their most efficacious application for a particular molecule to be purified. The purified product may be concentrated by filtration or ultrafiltration, and stored at a temperature at which the stability of the product is maximized.

There are a wide array of purification methods known to the art and the preceding method of purification is not meant to be limiting. Such purification techniques are described, for example, in Bailey, J.E. & Ollis, D.F. Biochemical Engineering Fundamentals, McGraw-Hill: New York (1986).

The identity and purity of the isolated compounds may be assessed by techniques standard in the art. These include high-performance liquid chromatography (HPLC), spectroscopic methods, staining methods, thin layer chromatography, NIRS, enzymatic assay, or microbiologically. Such analysis methods are reviewed in: Patek et al. (1994) *Appl Environ. Microbiol.* 60: 133-140; Malakhova et al. (1996) *Biotekhnologiya* 11: 27-32; and Schmidt et al. (1998) *Bioprocess Engineer.* 19: 67-70. Ulmann's Encyclopedia of Industrial Chemistry, (1996) vol. A27, VCH: Weinheim, p. 89-90, p. 521-540, p. 540-547, p. 559-566, 575-581 and p. 581-587; Michal, G. (1999) Biochemical Pathways: An Atlas of Biochemistry and Molecular Biology, John Wiley and Sons; Fallon, A. et al. (1987) Applications of HPLC in Biochemistry in: Laboratory Techniques in Biochemistry and Molecular Biology, vol. 17.

Equivalents

Those skilled in the art will recognize, or will be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. Such equivalents are intended to be encompassed by the following claims.

TABLE 1: GENES IN THE APPLICATION

Fatty acid and lipid synthesis

Identification Code	Contig.	NT Start	NT Stop	Gene Name	Function
RXA02335	GR00672	550	2322	EC-accC	BIOTIN CARBOXYLASE (EC 6 3 4 14)
RXA02173	GR00641	7473	8924	BS-yfiI,EC-accD	ACETYL-COENZYME A CARBOXYLASE CARBOXYL TRANSFERASE SUBUNIT BETA (EC 6 4 1 2)
RXA01764	GR00500	2178	3110	BS-fabG,EC-fabG	3-OXOACYL-[ACYL-CARRIER PROTEIN] REDUCTASE (EC 1 1.1.100)
RXA02487	GR00718	4937	4650		LONG-CHAIN-FATTY-ACID--COA LIGASE (EC 6.2 1 3)
RXA02490	GR00720	817	5		LONG-CHAIN-FATTY-ACID--COA LIGASE (EC 6.2 1.3)
RXA01467	GR00422	920	1210		ACYL CARRIER PROTEIN
RXA00798	GR00212	202	5		Acyl carrier protein phosphodiesterase
RXA01897	GR00544	617	1159		Acyl carrier protein phosphodiesterase
RXA02809	GR00790	277	5		Acyl carrier protein phosphodiesterase
RXA00113	GR00017	2	3295		FATTY-ACID SYNTHASE (EC 2 3 1 85)
RXA00158	GR00024	2088	4		FATTY ACID SYNTHASE (EC 2 3 1 85)
RXA00572	GR00155	2	3832		FATTY ACID SYNTHASE (EC 2 3 1 85)
RXA02582	GR00741	1890	6719	BS-pksP,EC-fabF	PROBABLE POLYKETIDE SYNTHASE CY338 20
RXA02691	GR00754	15347	14541	BS-yvoA,EC-farR	FATTY ACYL RESPONSIVE REGULATOR
RXA00880	GR00242	6213	8057		LONG-CHAIN-FATTY-ACID--COA LIGASE (EC 6.2 1 3)
RXA01060	GR00296	9566	10489		OMEGA-3 FATTY ACID DESATURASE (EC 1 14 99 -)
RXA01722	GR00488	5746	4022		MEDIUM-CHAIN-FATTY-ACID--COA LIGASE (EC 6 2 1 -)
RXA01644	GR00456	9854	8577	EC-cfa	CYCLOPROPANE-FATTY-ACYL-PHOSPHOLIPID SYNTHASE (EC 2.1.1.79)
RXA02029	GR00618	356	1669		CYCLOPROPANE-FATTY-ACYL-PHOSPHOLIPID SYNTHASE (EC 2.1.1 79)
RXA01801	GR00509	3396	2380		ENOYL-COA HYDRATASE (EC 4 2 1 17)
RXA02512	GR00721	3303	4259		LIPID A BIOSYNTHESIS LAUROYL ACYLTRANSFERASE (EC 2 3 1.-)
RXA00899	GR00245	1599	2864	EC-cls,BS-ywiE	CARDIOLIPIN SYNTHETASE (EC 2 7 8 -)
RXA00819	GR00221	18	1007	BS-yngJ,EC-caiA	ACYL-COA DEHYDROGENASE (EC 1.3.99 -)
RXA01766	GR00500	4081	4371		ACYL-COA DEHYDROGENASE (EC 1 3 99 -)
RXA01762	GR00500	1272	10	BS-yhfL,EC-fadD	LONG-CHAIN-FATTY-ACID--COA LIGASE (EC 6 2 1 3)
RXA00661	GR00179	3405	2662		3-OXOACYL-[ACYL-CARRIER PROTEIN] REDUCTASE (EC 1 1 1 100)
RXA00802	GR00214	3603	4516		3-OXOACYL-[ACYL-CARRIER PROTEIN] REDUCTASE (EC 1 1.1.100)
RXA02133	GR00639	3	308	EC-ydfG	3-OXOACYL-[ACYL-CARRIER PROTEIN] REDUCTASE (EC 1.1 1.100)
RXA01114	GR00308	2	793	EC-b1397,BS-yusK	3-KETOACYL-COA THIOLASE (EC 2 3 1.16)
RXA02810	GR00791	348	4		FATTY ACYL RESPONSIVE REGULATOR
RXA01894	GR00542	1622	2476	EC-cdsA,BS-cdsA	PHOSPHATIDATE CYTIDYLYLTRANSFERASE (EC 2 7.7.41)
RXA02599	GR00742	3179	3655		PHOSPHATIDYLGLYCEROPHOSPHATASE B (EC 3 1 3 27)
RXA02638	GR00749	8	511		1-ACYL-SN-GLYCEROL-3-PHOSPHATE ACYLTRANSFERASE (EC 2.3.1.51)

Identification Code	Contig.	NT Start	NT Stop	Gene Name	Function
RXA00856	GR00232	720	1256	BS-pgsA	CDP-DIACYLGLYCEROL--GLYCEROL-3-PHOSPHATE 3-PHOSPHATIDYLTRANSFERASE (EC 2 7 8 5)
RXA02511	GR00721	2621	3277		CDP-DIACYLGLYCEROL--GLYCEROL-3-PHOSPHATE 3-PHOSPHATIDYLTRANSFERASE (EC 2 7 8 5)
RXA02836	GR00827	106	411	BS-yqjO	KETOACYL REDUCTASE HETN (EC 1 3 1 -)
RXA02578	GR00740	2438	3541		PUTATIVE ACYLTRANSFERASE
RXA02150	GR00639	18858	19658		1-ACYL-SN-GLYCEROL-3-PHOSPHATE ACYLTRANSFERASE (EC 2.3.1.51)
RXA00607	GR00160	1869	2249		POLY(3-HYDROXYALKANOATE) POLYMERASE (EC 2.3.1.-)
RXA02397	GR00698	1688	2683		POLY-BETA-HYDROXYBUTYRATE POLYMERASE (EC 2 3 1 -)
RXA00660	GR00171	1027	5		HYDROXYACYLGLUTATHIONE HYDROLASE (EC 3 1.2.6)
RXA00801	GR00214	3138	3770		HYDROXYACYLGLUTATHIONE HYDROLASE (EC 3 1 2 6)
RXA00821	GR00221	1469	2311		HYDROXYACYLGLUTATHIONE HYDROLASE (EC 3 1 2 6)
RXA01833	GR00517	1666	260	BS-yrkH	HYDROXYACYLGLUTATHIONE HYDROLASE (EC 3 1 2 6)
RXA01853	GR00525	5561	5010	EC-b0927	HYDROXYACYLGLUTATHIONE HYDROLASE (EC 3 1 2 6)
RXA02424	GR00706	808	428		HYDROXYACYLGLUTATHIONE HYDROLASE (EC 3 1 2 6)
RXA00419	GR00095	3	464		ACETOACETYL-COA REDUCTASE (EC 1 1.1.36)
RXA00421	GR00096	565	723		ACETOACETYL-COA REDUCTASE (EC 1 1.1.36)

Polyketide Synthesis

Identification Code	Contig	NT Start	NT Stop	Gene Name	Function
RXA01420	GR00416	775	17		4"-MYCAROSYL ISOVALERYL-COA TRANSFERASE (EC 2.-.-.-)
RXA02581	GR00741	1	1527		POLYKETIDE SYNTHASE
RXA02582	GR00741	1890	6719	BS-pksP,EC-fabF	PROBABLE POLYKETIDE SYNTHASE CY338.20
RXA01138	GR00318	1656	2072		ACTINORHODIN POLYKETIDE DIMERASE (EC -.-.-.-)
RXA01980	GR00573	1470	838		POLYKETIDE CYCLASE

Fatty acid degradation

Identification Code	Contig.	NT Start	NT Stop	Gene Name	Function
RXA02268	GR00655	2182	3081		LIPASE (EC 3 1 1 3)
RXA02269	GR00655	3094	4065		LIPASE (EC 3.1.1.3)
RXA01614	GR00449	8219	7197		LYSOPHOSPHOLIPASE L2 (EC 3.1.1.5)
RXA01983	GR00573	3559	3053		LIPASE (EC 3.1.1.3)

Identification Code	Contig.	NT Start	NT Stop	Gene Name	Function
RXA02320	GR00667	593	6		PROPIONYL-COA CARBOXYLASE BETA CHAIN (EC 6 4 1 3)
RXA02321	GR00667	1380	937		PROPIONYL-COA CARBOXYLASE BETA CHAIN (EC 6 4 1 3)
RXA02343	GR00675	1403	1816		PROPIONYL-COA CARBOXYLASE BETA CHAIN (EC 6 4 1 3)
RXA02583	GR00741	6743	8290	BS-yqjD	PROPIONYL-COA CARBOXYLASE BETA CHAIN (EC 6 4 1 3)
RXA02850	GR00850	2	493		PROPIONYL-COA CARBOXYLASE BETA CHAIN (EC 6.4.1.3)
RXA02851	GR00851	524	6		PROPIONYL-COA CARBOXYLASE BETA CHAIN (EC 6.4.1.3)
RXA00870	GR00239	809	2320	BS-lolA	METHYLMALONATE-SEMIALDEHYDE DEHYDROGENASE (ACYLATING) (EC 1 2 1.27)
RXA01260	GR00367	2381	1200	BS-pdhD,EC-lpdA	2-Methyl-3-oxopropanoate:NAD+ oxidoreductase (CoA-propanoylating)
RXA01261	GR00367	2607	2437		LIPOAMIDE DEHYDROGENASE COMPONENT (E3) OF BRANCHED-CHAIN ALPHA-KETO ACID DEHYDROGENASE COMPLEX (EC 1.8.1.4)
RXA00931	GR00253	4959	4114	EC-tesB	LIPOAMIDE DEHYDROGENASE COMPONENT (E3) OF BRANCHED-CHAIN ALPHA-KETO ACID DEHYDROGENASE COMPLEX (EC 1 8 1 4)
RXA01136	GR00318	685	1116		Thioesterase II
RXA00559	GR00149	218	6		ISOVALERYL-COA DEHYDROGENASE (EC 1 3.99.10)
RXA01580	GR00440	707	6		PROTEIN VDLD
RXA02677	GR00754	3119	3877	BS-yhdW	Glycerophosphoryl diester phosphodiesterase
					GLYCEROPHOSPHORYL DIESTER PHOSPHODIESTERASE (EC 3 1 4 46)

Lipoprotein and Lipopolysaccharide synthesis

Identification Code	Contig.	NT Start	NT Stop	Gene Name	Function
RXA00002	GR00001	2278	1595		DOLICHOL-PHOSPHATE MANNOSYLTRANSFERASE (EC 2 4 1 83) / APOLIPOPROTEIN N-ACYLTRANSFERASE (EC 2 3 1 -)
RXA00160	GR00024	4044	4616		LIPOPROTEIN NLPD/LPPB HOMOLOG PRECURSOR
RXA00345	GR00064	90	1040		Zn-binding lipoprotein
RXA00413	GR00092	3859	2963	EC-yaeC,BS-yhcJ	OUTER MEMBRANE LIPOPROTEIN 3 PRECURSOR
RXA00482	GR00119	18691	18244		OUTER MEMBRANE LIPOPROTEIN BLC PRECURSOR
RXA01164	GR00332	1579	5	EC-lnt	DOLICHOL-PHOSPHATE MANNOSYLTRANSFERASE (EC 2 4 1 83) / APOLIPOPROTEIN N-ACYLTRANSFERASE (EC 2 3 1 -)
RXA01168	GR00333	1285	566		DOLICHOL-PHOSPHATE MANNOSYLTRANSFERASE (EC 2 4 1 83) / APOLIPOPROTEIN N-ACYLTRANSFERASE (EC 2 3 1 -)
RXA02082	GR00526	3159	1990		Lipopolysaccharide N-acetylglucosaminyltransferase
RXA02222	GR00551	9420	9794		PUTATIVE HOST CELL SURFACE-EXPOSED LIPOPROTEIN
RXA02313	GR00565	5812	4592		Lipopolysaccharide N-acetylglucosaminyltransferase
RXA02491	GR00720	902	2155	BS-ybbN	Lipopolysaccharide N-acetylglucosaminyltransferase
RXA02595	GR00741	19052	19702		Lipopolysaccharide N-acetylglucosaminyltransferase
RXA02616	GR00745	598	1308	EC-nlpD	LIPOPROTEIN NLPD PRECURSOR
RXA02627	GR00747	2981	2139		DTXR/IRON-REGULATED LIPOPROTEIN PRECURSOR
RXA02650	GR00752	1460	2038		LIPOPROTEIN SIGNAL PEPTIDASE (EC 3 4 23 36)

Identification Code	Contig	NT Start	NT Stop	Gene Name	Function
RXA01094	GR00306	2703	1756		PROLIPOPROTEIN DIACYLGLYCERYL TRANSFERASE (EC 2.4.99.-)
RXA01985	GR00574	466	1008		DTXR/IRON-REGULATED LIPOPROTEIN PRECURSOR
RXA02804	GR00785	1	261		DTXR/IRON-REGULATED LIPOPROTEIN PRECURSOR
RXA00934	GR00253	6835	6047		(AE000805) LPS biosynthesis RfbU related protein [Methanobacterium thermoautotrophicum]
RXA02605	GR00742	11557	12051		ANTIGEN 85-B PRECURSOR

Terpenoid biosynthesis

Identification Code	Contig	NT Start	NT Stop	Gene Name	Function
RXA00875	GR00241	2423	1857	EC-b2889	ISOPENTENYL-DIPHOSPHATE DELTA-ISOMERASE (EC 5.3.3.2)
RXA01292	GR00373	1204	2388		PHYTOENE DEHYDROGENASE (EC 1.3.-.-)
RXA01293	GR00373	2370	2696		PHYTOENE DEHYDROGENASE (EC 1.3.-.-)
RXA02310	GR00665	1132	2394		GERANYLGERANYL HYDROGENASE
RXA02716	GR00758	18539	19585		GERANYLGERANYL PYROPHOSPHATE SYNTHASE (EC 2.5.1.1)
RXA01067	GR00298	1453	2181	EC-b0174	undecaprenyl-diphosphate synthase (EC 2.5.1.31)
RXA01269	GR00367	20334	19894	BS-yvtC,EC-b2047	UNDECAPRENYL-PHOSPHATE GALACTOSEPHOSPHOTRANSFERASE (EC 2.7.8.6)
RXA01205 (EC 2.4.1.-)	GR00346	3		533	PUTATIVE UNDECAPRENYL-PHOSPHATE ALPHA-N-ACETYLGLUCOSAMINYLTRANSFERASE
RXA01576	GR00438	8053	8811		DOLICHYL-PHOSPHATE BETA-GLUCOSYLTRANSFERASE (EC 2.4.1.117)
RXA01258	GR00366	1150	2709		DOLICHYL-PHOSPHATE-MANNOSE--PROTEIN MANNOSYLTRANSFERASE 1 (EC 2.4.1.109)
RXA01351	GR00392	2841	3137	BS-yvdD	(U15180) P450 cytochrome,isopentenyltransf, ferridox [Mycobacterium leprae]
RXA02309	GR00665	978	4	BS-gerCC,EC-ispB	OCTAPRENYL-DIPHOSPHATE SYNTHASE (EC 2.5.1.-)
RXA00477	GR00119	13187	11544		PHYTOENE DEHYDROGENASE (EC 1.3.-.-)
RXA00478	GR00119	14020	13190		PHYTOENE SYNTHASE (EC 2.5.1.-)
RXA01291	GR00373	345	1277		PHYTOENE SYNTHASE (EC 2.5.1.-)
RXA00480	GR00119	17444	16329		FARNESYL DIPHOSPHATE SYNTHASE (EC 2.5.1.1) (EC 2.5.1.10)

ABC-Transporter

Identification Code	Contig	NT Start	NT Stop	Gene Name	Function
RXA01946	GR00559	1649	575		(AL021184) ABC transporter ATP binding protein [Mycobacterium tuberculosis]
RXA00164	GR00025	1782	94	EC-msbA	, P, G, R ATPase subunits of ABC transporters
RXA00165	GR00025	3275	1860		, P, G, R ATPase subunits of ABC transporters
RXA00243	GR00037	930	4		, P, G, R ATPase subunits of ABC transporters

Identification Code	Contig	NT Start	NT Stop	Gene Name	Function
RXA00259	GR00039	6469	6268		, P, G, R ATPase subunits of ABC transporters
RXA00410	GR00092	629	164		, P, G, R ATPase subunits of ABC transporters
RXA00456	GR00114	316	5		, P, G, R ATPase subunits of ABC transporters
RXA00459	GR00115	1231	245		, P, G, R ATPase subunits of ABC transporters
RXA01604	GR00448	2	607		, P, G, R ATPase subunits of ABC transporters
RXA02547	GR00726	22055	19932		, P, G, R ATPase subunits of ABC transporters
RXA02571	GR00736	1469	2497		, P, G, R ATPase subunits of ABC transporters
RXA02074	GR00628	5798	4176		, P, G, R ATPase subunits of ABC transporters
RXA02095	GR00629	14071	15474		, P, G, R ATPase subunits of ABC transporters
RXA02225	GR00652	3156	2275		, P, G, R ATPase subunits of ABC transporters
RXA02253	GR00654	20480	21406	BS-yvcJ,EC-b3205	, P, G, R ATPase subunits of ABC transporters
RXA01881	GR00537	3092	3532		ATPase components of ABC transporters with duplicated ATPase domains
RXA00188	GR00029	1447	5	EC-yjjK,BS-ytmR	Hypothetical ABC Transporter ATP-Binding Protein
RXA00526	GR00136	1353	664	BS-yknY,EC-b0879	Hypothetical ABC Transporter ATP-Binding Protein
RXA00733	GR00197	411	4		Hypothetical ABC Transporter ATP-Binding Protein
RXA00734	GR00197	863	411		Hypothetical ABC Transporter ATP-Binding Protein
RXA00735	GR00198	849	181		Hypothetical ABC Transporter ATP-Binding Protein
RXA00878	GR00242	3733	1871	BS-ywjA	Hypothetical ABC Transporter ATP-Binding Protein
RXA01191	GR00341	1571	165	EC-mdlB	Hypothetical ABC Transporter ATP-Binding Protein
RXA01212	GR00350	1	813	EC-yadG	Hypothetical ABC Transporter ATP-Binding Protein
RXA01947	GR00560	812	534		Hypothetical ABC Transporter ATP-Binding Protein
RXA02749	GR00764	4153	5028	BS-yfiL	Hypothetical ABC Transporter ATP-Binding Protein
RXA02224	GR00652	2271	475		Hypothetical ABC Transporter ATP-Binding Protein
RXA00525	GR00136	664	5		Hypothetical ABC Transporter Permease Protein
RXA00556	GR00146	1	594	BS-yknZ	Hypothetical ABC Transporter Permease Protein
RXA02750	GR00764	5079	5894		Hypothetical ABC Transporter Permease Protein
RXA02096	GR00629	15458	16774		Hypothetical ABC Transporter Permease Protein
RXA01190	GR00340	1162	194		MULTIDRUG RESISTANCE-LIKE ATP-BINDING PROTEIN MDL
RXA01808	GR00509	8993	7875		PUTATIVE ABC TRANPORTER
RXA02562	GR00732	796	1515		PUTATIVE ABC TRANSPORTER
RXA00950	GR00260	173	1078	BS-ydbJ	similar to ABC transporter (ATP-binding protein) START CODON GTG
RXA02119	GR00636	4222	2582	BS-ydiF	similar to ABC transporter (ATP-binding protein)
RXA00431	GR00099	119	793	BS-tagH	ABCA PROTEIN two-component ABC transporter involved in the metabolism of two wall teichoic acids
RXA01185	GR00338	2451	1594		ATP-BINDING PROTEIN
RXA02434	GR00708	4642	3590		ATP-BINDING PROTEIN
RXA00412	GR00092	2764	1685	BS-yusC,EC-abc	ATP-BINDING PROTEIN ABC
RXA01323	GR00365	1175	3439	BS-yvgX,EC-b0484	similar to heavy metal-transporting ATPase
RXA00001	GR00001	1386	259	BS-msmX,EC-ugpC	SN-GLYCEROL-3-PHOSPHATE TRANSPORT ATP-BINDING PROTEIN UGPC

Other transporters

Identification Code	Contig.	NT Start	NT Stop	Gene Name	Function
RXA02402	GR00700	747	4		(AF027868) putative transporter [Bacillus subtilis]
RXA02626	GR00747	2012	1008	BS-yocS	(AF027868) putative transporter [Bacillus subtilis]
RXA02261	GR00654	30936	32291	BS-nrgA,EC-amtB	AMMONIUM TRANSPORT SYSTEM
RXA02020	GR00613	1015	5	BS-gabP,EC-gabP	AROMATIC AMINO ACID TRANSPORT PROTEIN AROP
RXA00281	GR00043	4721	5404		BACITRACIN TRANSPORT ATP-BINDING PROTEIN BCRA
RXA00570	GR00153	1	498		BENZOATE MEMBRANE TRANSPORT PROTEIN
RXA00571	GR00154	2	1186		BENZOATE MEMBRANE TRANSPORT PROTEIN
RXA00962	GR00268	2	667		BENZOATE MEMBRANE TRANSPORT PROTEIN
RXA02811	GR00792	177	560		BENZOATE MEMBRANE TRANSPORT PROTEIN
RXA02115	GR00635	2	1198		BENZOATE MEMBRANE TRANSPORT PROTEIN
RXA00590	GR00157	178	564		BRANCHED CHAIN AMINO ACID TRANSPORT SYSTEM II CARRIER PROTEIN
RXA01538	GR00427	5040	5429	BS-ycsG	BRANCHED CHAIN AMINO ACID TRANSPORT SYSTEM II CARRIER PROTEIN
RXA01727	GR00489	1471	194	BS-brnQ,EC-brnQ	BRANCHED-CHAIN AMINO ACID TRANSPORT SYSTEM CARRIER PROTEIN
RXA00823	GR00163	6525	7862	EC-dctA,BS-ydbH	C4-DICARBOXYLATE TRANSPORT PROTEIN
RXA01584	GR00441	55	597		CHROMATE TRANSPORT PROTEIN
RXA00702	GR00182	2165	846	BS-ykoD	COBALT TRANSPORT ATP-BINDING PROTEIN CBIO
RXA00828	GR00223	1687	1319		COBALT TRANSPORT ATP-BINDING PROTEIN CBIO
RXA00852	GR00231	3137	2448	BS-ybxA	COBALT TRANSPORT ATP-BINDING PROTEIN CBIO
RXA00690	GR00181	1213	68		COBALT TRANSPORT PROTEIN CBIO
RXA00827	GR00223	1319	567		COBALT TRANSPORT PROTEIN CBIQ
RXA00851	GR00231	2448	1840		COBALT TRANSPORT PROTEIN CBIQ
RXA00939	GR00256	1501	1334		COPPER/POTASSIUM-TRANSPORTING ATPASE B (EC 3.6.1.36)
RXA01245	GR00360	2	1768		COPPER/POTASSIUM-TRANSPORTING ATPASE B (EC 3.6.1.36)
RXA01247	GR00361	256	489		COPPER/POTASSIUM-TRANSPORTING ATPASE B (EC 3.6.1.36)
RXA00099	GR00014	8172	9344	BS-yycB	CYANATE TRANSPORT PROTEIN CYNX
RXA01584	GR00437	746	9	EC-b1791	CYANATE TRANSPORT PROTEIN CYNX
RXA00634	GR00166	3732	5114	EC-b0709	DI-/TRIPEPTIDE TRANSPORTER
RXA02451	GR00710	3484	5007	BS-ydtF,EC-b1634	DI-/TRIPEPTIDE TRANSPORTER
RXA02394	GR00697	1895	585	EC-yjaN	DICARBOXYLATE TRANSPORTER
RXA01012	GR00288	3748	2108	EC-b0829,BS-appF	DIPEPTIDE TRANSPORT ATP-BINDING PROTEIN DPPD
RXA02660	GR00753	548	1186		DIPEPTIDE TRANSPORT SYSTEM PERMEASE PROTEIN DPPB
RXA02661	GR00753	1239	1457		DIPEPTIDE TRANSPORT SYSTEM PERMEASE PROTEIN DPPB
RXA02034	GR00619	1787	822	BS-appB	DIPEPTIDE TRANSPORT SYSTEM PERMEASE PROTEIN DPPB
RXA01013	GR00288	4549	3755		DIPEPTIDE TRANSPORT SYSTEM PERMEASE PROTEIN DPPC
RXA02033	GR00619	800	12		DIPEPTIDE TRANSPORT SYSTEM PERMEASE PROTEIN DPPC
RXA01006	GR00287	862	5		DIPEPTIDE TRANSPORTER PROTEIN DPPB
RXA02312	GR00665	4459	3101	EC-cycA,BS-ydgF	D-SERINE/D-ALANINE/GLYCINE TRANSPORTER
RXA02762	GR00768	346	630		D-XYLOSE PROTON-SYMPORTER
RXA00090	GR00013	6644	7762	BS-ycjO	FERRIC ANGUIBACTIN TRANSPORT SYSTEM PERMEASE PROTEIN FATC
RXA00089	GR00013	5656	6654	BS-ycjN	FERRIC ANGUIBACTIN TRANSPORT SYSTEM PERMEASE PROTEIN FATD
RXA01285	GR00371	3	545		FERRIC ENTEROBACTIN TRANSPORT ATP-BINDING PROTEIN FEPC
RXA02728	GR00761	184	996		FERRIC ENTEROBACTIN TRANSPORT ATP-BINDING PROTEIN FEPC

Identification Code	Contig	NT Start	NT Stop	Gene Name	Function
RXA02864	GR10007	2808	2027		FERRIC ENTEROBACTIN TRANSPORT ATP-BINDING PROTEIN FEPC
RXA00523	GR00135	30	779		FERRIC ENTEROBACTIN TRANSPORT PROTEIN FEPG
RXA01289	GR00372	2376	3419	EC-fepD	FERRIC ENTEROBACTIN TRANSPORT PROTEIN FEPG
RXA01290	GR00372	3412	4575	EC-fepG,BS-yfhA	FERRIC ENTEROBACTIN TRANSPORT PROTEIN FEPG
RXA01822	GR00516	6	587		FERRIC ENTEROBACTIN TRANSPORT PROTEIN FEPG
RXA00466	GR00117	947	1933		Ferrichrome transport proteins
RXA02863	GR10007	2000	1026		Ferrichrome transport proteins
RXA01986	GR00575	622	5		GALACTOSE-PROTON SYMPORT
RXA02447	GR00710	1	270		GALACTOSE-PROTON SYMPORT
RXA02761	GR00768	153	353		GALACTOSE-PROTON SYMPORT
RXA02769	GR00771	1	711	BS-ydjK,EC-xylE	GALACTOSE-PROTON SYMPORT
RXA00092	GR00014	1	204		GLUTAMINE TRANSPORT ATP-BINDING PROTEIN GLND
RXA02818	GR00745	1914	2351		GLUTAMINE TRANSPORT ATP-BINDING PROTEIN GLND
RXA02900	GR10040	2979	2128	BS-ytsC	GLUTAMINE TRANSPORT ATP-BINDING PROTEIN GLND
RXA01591	GR00446	3	947		GLYCINE BETAINE TRANSPORTER BETP
RXA00201	GR00032	191	6		HIGH AFFINITY RIBOSE TRANSPORT PROTEIN RBSD
RXA01221	GR00354	2108	2833	EC-livG	HIGH-AFFINITY BRANCHED-CHAIN AMINO ACID TRANSPORT ATP-BINDING PROTEIN BRAG
RXA01222	GR00354	2844	3542	EC-livF	HIGH-AFFINITY BRANCHED-CHAIN AMINO ACID TRANSPORT ATP-BINDING PROTEIN LIVF
RXA01219	GR00354	151	1032		HIGH-AFFINITY BRANCHED-CHAIN AMINO ACID TRANSPORT PERMEASE PROTEIN LIVH
RXA01220	GR00354	1032	2108		HIGH-AFFINITY BRANCHED-CHAIN AMINO ACID TRANSPORT PERMEASE PROTEIN LIVM
RXA00091	GR00013	7762	6514	BS-yclP	IRON(III) DICITRATE TRANSPORT ATP-BINDING PROTEIN FECE
RXA00228	GR00032	29232	28642		IRON(III) DICITRATE TRANSPORT ATP-BINDING PROTEIN FECE
RXA00346	GR00064	1054	1743		IRON(III) DICITRATE TRANSPORT ATP-BINDING PROTEIN FECE
RXA00524	GR00135	779	1111		IRON(III) DICITRATE TRANSPORT ATP-BINDING PROTEIN FECE
RXA01823	GR00516	591	1367	BS-yusV,EC-fepC	IRON(III) DICITRATE TRANSPORT ATP-BINDING PROTEIN FECE
RXA02767	GR00770	1032	1814	BS-ylmA	IRON(III) DICITRATE TRANSPORT ATP-BINDING PROTEIN FECE
RXA02792	GR00777	8581	7829		IRON(III) DICITRATE TRANSPORT ATP-BINDING PROTEIN FECE
RXA01235	GR00358	1165	194	BS-yfmD	IRON(III) DICITRATE TRANSPORT SYSTEM PERMEASE PROTEIN FECD
RXA01419	GR00415	686	1151		IRON(III) DICITRATE TRANSPORT SYSTEM PERMEASE PROTEIN FECD
RXA02794	GR00777	10172	9552	EC-fecC	IRON(III) DICITRATE TRANSPORT SYSTEM PERMEASE PROTEIN FECD
RXA02865 PROTEIN AROP	GR10007	3832	2816		IRON(III) DICITRATE TRANSPORT SYSTEM PERMEASE PROTEIN FECD MATIC AMINO ACID TRANSPORT
RXA00123	GR00019	7029	5911		MAGNESIUM AND COBALT TRANSPORT PROTEIN CORA
RXA02354	GR00682	473	1261		MALTOSE TRANSPORT SYSTEM PERMEASE PROTEIN MALG
RXA02441	GR00709	5940	5284		MANGANESE TRANSPORT SYSTEM ATP-BINDING PROTEIN MNTA
RXA02442	GR00709	5970	6818	EC-yebl	MANGANESE TRANSPORT SYSTEM MEMBRANE PROTEIN MNTB
RXA01756	GR00498	2069	762	BS-ykoK	MG2+ TRANSPORTER MGTE
RXA02068	GR00627	2	1120		MG2+ TRANSPORTER MGTE
RXA00665	GR00174	135	572	BS-yxiD	MG2+/CITRATE COMPLEX SECONDARY TRANSPORTER
RXA02808	GR00789	1	258		MG2+/CITRATE COMPLEX SECONDARY TRANSPORTER
RXA00444	GR00106	626	1402	BS-yvgM,EC-cysU	MOLYBDENUM TRANSPORT SYSTEM PERMEASE PROTEIN MODB
RXA02614	GR00743	5964	5236	BS-ygaL,EC-ycbE	NITRATE TRANSPORT ATP-BINDING PROTEIN NRTC
RXA01142	GR00320	721	302		NITRATE TRANSPORT ATP-BINDING PROTEIN NRTD
RXA01135	GR00318	327	4		NITRATE TRANSPORT PROTEIN NRTA
RXA01141	GR00319	636	175		NITRATE TRANSPORT PROTEIN NRTA

Identification Code	Contig.	NT Start	NT Stop	Gene Name	Function
RXA00728	GR00193	1658	2449	BS-yckA,EC-hisM	NOPALINE TRANSPORT SYSTEM PERMEASE PROTEIN NOCM
RXA02795	GR00778	3	1097		OLIGOPEPTIDE TRANSPORT ATP-BINDING PROTEIN APPF
RXA00761	GR00203	8530	9120		OLIGOPEPTIDE TRANSPORT ATP-BINDING PROTEIN OPPD
RXA01939	GR00556	2042	1440		OLIGOPEPTIDE TRANSPORT ATP-BINDING PROTEIN OPPD
RXA02863	GR00753	2059	3453	BS-oppD,EC-oppD	OLIGOPEPTIDE TRANSPORT ATP-BINDING PROTEIN OPPD
RXA02864	GR00753	3611	4270		OLIGOPEPTIDE TRANSPORT ATP-BINDING PROTEIN OPPF
RXA00759	GR00203	6580	7503	EC-dppB,BS-dppB	OLIGOPEPTIDE TRANSPORT SYSTEM PERMEASE PROTEIN OPPB
RXA00760	GR00203	7499	8530	BS-appC,EC-b0832	OLIGOPEPTIDE TRANSPORT SYSTEM PERMEASE PROTEIN OPPC
RXA02035	GR00619	3295	1787	EC-b0830	PERIPLASMIC DIPEPTIDE TRANSPORT PROTEIN PRECURSOR
RXA00775	GR00205	6057	5287	EC-pstB,BS-yqgK	PHOSPHATE TRANSPORT ATP-BINDING PROTEIN PSTB
RXA00776	GR00205	7016	6096	EC-pstA,BS-yqgI	PHOSPHATE TRANSPORT SYSTEM PERMEASE PROTEIN PSTA
RXA00777	GR00205	8098	7034	EC-pstC,BS-yqgH	PHOSPHATE TRANSPORT SYSTEM PERMEASE PROTEIN PSTC
RXA00774	GR00205	4546	5199		PHOSPHATE TRANSPORT SYSTEM REGULATORY PROTEIN
RXA01002	GR00285	3	419	EC-phnC	PHOSPHONATES TRANSPORT ATP-BINDING PROTEIN PHNC
RXA01000	GR00284	2	541		PHOSPHONATES TRANSPORT SYSTEM PERMEASE PROTEIN PHNE
RXA01003	GR00285	419	1222		PHOSPHONATES TRANSPORT SYSTEM PERMEASE PROTEIN PHNE
RXA00193	GR00029	10101	9259		POTENTIAL STARCH DEGRADATION PRODUCTS TRANSPORT SYSTEM PERMEASE PROTEIN AMYD
RXA01298	GR00374	1254	882	EC-ugpA	POTENTIAL STARCH DEGRADATION PRODUCTS TRANSPORT SYSTEM PERMEASE PROTEIN AMYD
RXA02422	GR00705	8200	8634		POTENTIAL STARCH DEGRADATION PRODUCTS TRANSPORT SYSTEM PERMEASE PROTEIN AMYD
RXA02515	GR00723	984	1719	BS-yuiY,EC-b1682	PROBABLE ATP-DEPENDENT TRANSPORTER YCF16
RXA00732	GR00196	826	5	BS-yfiB	PROBABLE TRANSPORT ATP-BINDING PROTEIN MSBA
RXA00181	GR00028	3954	2383	BS-opuE,EC-putP	PROLINE TRANSPORT SYSTEM
RXA00591	GR00158	229	1581	EC-b1981	PROLINE/BETAINE TRANSPORTER
RXA01629	GR00453	3476	1965	EC-proP	PROLINE/BETAINE TRANSPORTER
RXA02030	GR00618	3072	1687	EC-yhjE	PROLINE/BETAINE TRANSPORTER
RXA01995	GR00584	1362	2015		PUTATIVE 3-(3-HYDROXYPHENYL) PROPIONATE TRANSPORT PROTEIN
RXA00953	GR00262	1830	1078	EC-yjiS	PUTATIVE TRANSPORT PROTEIN SGAT
RXA01188	GR00339	1585	462		PUTATIVE TRANSPORT PROTEIN SGAT
RXA01972	GR00569	2116	1523		QUATERNARY AMINE TRANSPORTER
RXA00204	GR00032	3783	2212	BS-rbsA,EC-rbsA	RIBOSE TRANSPORT ATP-BINDING PROTEIN RBSA
RXA02438	GR00709	3236	2478		RIBOSE TRANSPORT ATP-BINDING PROTEIN RBSA
RXA00203	GR00032	2152	1241	BS-rbsC,EC-rbsC	RIBOSE TRANSPORT SYSTEM PERMEASE PROTEIN RBSC
RXA00270	GR00041	2720	1833		RIBOSE TRANSPORT SYSTEM PERMEASE PROTEIN RBSC
RXA02439	GR00709	4258	3236		RIBOSE TRANSPORT SYSTEM PERMEASE PROTEIN RBSC
RXA00311	GR00053	1592	736		SHIKIMATE TRANSPORTER
RXA00312	GR00053	2086	1641		SHIKIMATE TRANSPORTER
RXA01411	GR00412	1	327		SHIKIMATE TRANSPORTER
RXA01900	GR00544	2822	4120		SHIKIMATE TRANSPORTER
RXA02507	GR00720	19760	21160		SHIKIMATE TRANSPORTER
RXA00186	GR00028	12242	12988		SHORT-CHAIN FATTY ACIDS TRANSPORTER
RXA00187	GR00028	13097	13447	EC-aloB	SHORT-CHAIN FATTY ACIDS TRANSPORTER
RXA00445	GR00107	21	932		SN-GLYCEROL-3-PHOSPHATE TRANSPORT ATP-BINDING PROTEIN UGPC
RXA01890	GR00541	874	155		SN-GLYCEROL-3-PHOSPHATE TRANSPORT ATP-BINDING PROTEIN UGPC
RXA02353	GR00682	6	473		SN-GLYCEROL-3-PHOSPHATE TRANSPORT SYSTEM PERMEASE PROTEIN UGPA
RXA01297	GR00374	826	29	BS-yurM,EC-ugpE	SN-GLYCEROL-3-PHOSPHATE TRANSPORT SYSTEM PERMEASE PROTEIN UGPE

Identification Code	Contig	NT Start	NT Stop	Gene Name	Function
RXA01667	GR00464	703	1908		SODIUM/GLUTAMATE SYMPORT CARRIER PROTEIN
RXA02171	GR00641	6571	4919	BS-ywcA,EC-yjcG	SODIUM/PROLINE SYMPORTER
RXA00902	GR00245	4643	5875		SODIUM-DEPENDENT PHOSPHATE TRANSPORT PROTEIN
RXA00941	GR00257	1999	683	BS-ycel,EC-ydjE	sodium-dependent phosphate transport protein
RXA00449	GR00109	2040	1036		Sodium-Dicarboxylate Symport Protein
RXA01755	GR00498	352	5		Sodium-Dicarboxylate Symport Protein
RXA00269	GR00041	1826	1038		SPERMIDINE/PUTRESCINE TRANSPORT ATP-BINDING PROTEIN POTA
RXA00368	GR00076	1	579		SPERMIDINE/PUTRESCINE TRANSPORT ATP-BINDING PROTEIN POTA
RXA00369	GR00076	583	1299		SPERMIDINE/PUTRESCINE TRANSPORT ATP-BINDING PROTEIN POTA
RXA00370	GR00077	6	803		SPERMIDINE/PUTRESCINE TRANSPORT ATP-BINDING PROTEIN POTA
RXA02073	GR00628	4176	2647		TRANSPORT ATP-BINDING PROTEIN CYDC
RXA01399	GR00409	1	1119	BS-cydC	TRANSPORT ATP-BINDING PROTEIN CYDD
RXA01339	GR00389	8408	7164		TYROSINE-SPECIFIC TRANSPORT PROTEIN
RXA02527	GR00725	5519	6847	BS-yflS,EC-b0770	2-OXOGLUTARATE/MALATE TRANSLOCATOR PRECURSOR
RXA00298	GR00046	4459	6303	EC-betT,BS-opuD	Ectoine/Proline/Glycine betaine carrier ectP
RXA00586	GR00159	335	787		potassium efflux system protein phaE
RXA02044	GR00623	528	1910	BS-ykaB	similar to low-affinity inorganic phosphate transporter
RXA00104	GR00014	15895	16650		CYSQ PROTEIN, ammonium transport protein
RXA02364	GR00686	841	215	EC-yiaO	C4-DICARBOXYLATE-BINDING PERIPLASMIC PROTEIN PRECURSOR, transport protein
RXA02099	GR00630	6198	6470		AMMONIUM TRANSPORT SYSTEM

Permeases

Identification Code	Contig	NT Start	NT Stop	Gene Name	Function
RXA02561	GR00732	664	5	BS-ydeG	NUCLEOSIDE PERMEASE NUPG
RXA02566	GR00733	782	345		NUCLEOSIDE PERMEASE NUPG
RXA00034	GR00004	2	511		Permeases
RXA00190	GR00029	4544	2394		Permeases
RXA00842	GR00228	3208	2009	BS-ykvI	Permeases
RXA01553	GR00432	234	1553		Permeases
RXA00051	GR00008	5770	7173	EC-b0402,BS-ytnA	PROLINE-SPECIFIC PERMEASE PROY
RXA01172	GR00334	2687	4141	BS-ybaR	SULFATE PERMEASE
RXA02128	GR00837	2906	4600		SULFATE PERMEASE
RXA02634	GR00748	6045	7655	BS-yvdB,EC-ychM	SULFATE PERMEASE
RXA02233	GR00653	6856	6067	EC-b1006,BS-pyrP	URACIL PERMEASE
RXA02372	GR00688	6	560		XANTHINE PERMEASE
RXA02377	GR00689	3336	4526	EC-b2888,BS-pbuX	XANTHINE PERMEASE
RXA00081	GR00012	8453	7101		(Predicted) amino acid permeases

Identification Code	Contig	NT Start	NT Stop	Gene Name	Function
RXA02676	GR00754	2697	1309	EC-yjgT,BS-gnlP	GLUCONATE PERMEASE
RXA00432	GR00100	1	891	BS-alsT	NA(+)-LINKED D-ALANINE GLYCINE PERMEASE
RXA00436	GR00101	45	569	EC-yaaJ	NA(+)-LINKED D-ALANINE GLYCINE PERMEASE
RXA00847	GR00230	1829	381		OLIGOPEPTIDE-BINDING PROTEIN APPA PRECURSOR (permease)
RXA01382	GR00405	1067	6		OLIGOPEPTIDE-BINDING PROTEIN OPPA PRECURSOR (permease)
RXA02659	GR00753	2	313		OLIGOPEPTIDE-BINDING PROTEIN OPPA PRECURSOR (permease)

Channel Proteins

Identification Code	Contig	NT Start	NT Stop	Gene Name	Function
RXA01395	GR00408	6106	5021	EC-b3001	POTASSIUM CHANNEL BETA SUBUNIT
RXA01737	GR00493	2913	3971		POTASSIUM CHANNEL PROTEIN
RXA00596	GR00159	335	787		potassium efflux system protein phaE
RXA02079	GR00628	9034	9848		CATION EFFLUX SYSTEM PROTEIN CZCD
RXA01303	GR00376	1724	390	EC-narU	NITRITE EXTRUSION PROTEIN
RXA02079	GR00628	9034	9848		CATION EFFLUX SYSTEM PROTEIN CZCD
RXA00832	GR00224	2239	1685	EC-chaA	CALCIUM/PROTON ANTIPORTER
RXA00378	GR00081	3271	1499		Cation transport ATPases
RXA00942	GR00257	2408	2203		CATION-TRANSPORTING ATPASE PACS (EC 3.6.1.-)
RXA01338	GR00389	6984	5087		CATION-TRANSPORTING ATPASE PACS (EC 3.6.1.-)
RXA01625	GR00452	3850	3850		CATION-TRANSPORTING ATPASE PACS (EC 3.6.1.-)
RXA02220	GR00651	3205	5880	BS-yloB,EC-mgtA	CATION-TRANSPORTING ATPASE PMA1 (EC 3.6.1.-)
RXA00980	GR00276	2648	3286		CATION-TRANSPORTING P-TYPE ATPASE B (EC 3.6.1.-)
RXA02344	GR00676	682	5	EC-kup	KUP SYSTEM POTASSIUM UPTAKE PROTEIN
RXA02348	GR00677	1719	586		KUP SYSTEM POTASSIUM UPTAKE PROTEIN
RXA02426	GR00707	2165	633		PROBABLE NA(+)/H(+) ANTIPORTER
RXA00980	GR00266	563	105		PROTON/SODIUM-GLUTAMATE SYMPORT PROTEIN
RXA01070	GR00299	2089	704	EC-b1729	PROTON/SODIUM-GLUTAMATE SYMPORT PROTEIN
RXA02628	GR00748	6	410		LARGE CONDUCTANCE MECHANOSENSITIVE CHANNEL
RXA01395	GR00408	6106	5021	EC-b3001	POTASSIUM CHANNEL BETA SUBUNIT
RXA01737	GR00493	2913	3971		POTASSIUM CHANNEL PROTEIN

Other membrane proteins

Identification Code	Contig	NT Start	NT Stop	Gene Name	Function
RXA02597	GR00742	2329	542		OUTER MEMBRANE USHER PROTEIN FIMC PRECURSOR
RXA01454	GR00420	270	4		integral membrane protein
RXA01455	GR00420	745	284		integral membrane protein
RXA02684	GR00754	8923	8060	BS-ywlF	MEMBRANE-BOUND PROTEIN LYTR

TABLE 2: GENES IDENTIFIED FROM GENBANK

GenBank™ Accession No.	Gene Name	Gene Function	Reference
A09073	ppg	Phosphoenol pyruvate carboxylase	Bachmann, B. et al "DNA fragment coding for phosphoenolpyruvat corboxylase, recombinant DNA carrying said fragment, strains carrying the recombinant DNA and method for producing L-aminino acids using said strains," Patent: EP 0358940-A 3 03/21/90
A45579, A45581, A45583, A45585 A45587		Threonine dehydratase	Moeckel, B. et al. "Production of L-isoleucine by means of recombinant micro-organisms with deregulated threonine dehydratase," Patent. WO 9519442-A 5 07/20/95
AB003132	murC, ftsQ; ftsZ.		Kobayashi, M. et al. "Cloning, sequencing, and characterization of the ftsZ gene from coryneform bacteria," *Biochem. Biophys. Res. Commun.*, 236(2):383-388 (1997)
AB015023	murC; ftsQ		Wachi, M. et al. "A murC gene from Coryneform bacteria," *Appl. Microbiol. Biotechnol*, 51(2):223-228 (1999)
AB018530	dtsR		Kimura, E. et al. "Molecular cloning of a novel gene, dtsR, which rescues the detergent sensitivity of a mutant derived from *Brevibacterium lactofermentum*," *Biosci. Biotechnol. Biochem*, 60(10):1565-1570 (1996)
AB018531	dtsR1; dtsR2		
AB020624	murI	D-glutamate racemase	
AB023377	tkt	transketolase	
AB024708	gltB, gltD	Glutamine 2-oxoglutarate aminotransferase large and small subunits	
AB025424	acn	aconitase	
AB027714	rep	Replication protein	
AB027715	rep; aad	Replication protein; aminoglycoside adenyltransferase	
AF005242	argC	N-acetylglutamate-5-semialdehyde dehydrogenase	
AF005635	glnA	Glutamine synthetase	
AF030405	hisF	cyclase	
AF030520	argG	Argininosuccinate synthetase	
AF031518	argF	Ornithine carbamolytransferase	
AF036932	aroD	3-dehydroquinate dehydratase	

GenBank™ Accession No.	Gene Name	Gene Function	Reference
AF038548	pyc	Pyruvate carboxylase	
AF038651	dciAE; apt; rel	Dipeptide-binding protein; adenine phosphoribosyltransferase; GTP pyrophosphokinase	Wehmeier, L. et al. "The role of the Corynebacterium glutamicum rel gene in (p)ppGpp metabolism," *Microbiology*, 144.1853-1862 (1998)
AF041436	argR	Arginine repressor	
AF045998	impA	Inositol monophosphate phosphatase	
AF048764	argH	Argininosuccinate lyase	
AF049897	argC; argJ; argB; argD; argF; argR; argG; argH	N-acetylglutamylphosphate reductase, ornithine acetyltransferase; N-acetylglutamate kinase, acetylornithine transminase; ornithine carbamoyltransferase; arginine repressor; argininosuccinate synthase; argininosuccinate lyase	
AF050109	inhA	Enoyl-acyl carrier protein reductase	
AF050166	hisG	ATP phosphoribosyltransferase	
AF051846	hisA	Phosphoribosylformimino-5-amino-1-phosphoribosyl-4-imidazolecarboxamide isomerase	
AF052652	metA	Homoserine O-acetyltransferase	Park, S. et al. "Isolation and analysis of metA, a methionine biosynthetic gene encoding homoserine acetyltransferase in Corynebacterium glutamicum," *Mol Cells.*, 8(3):286-294 (1998)
AF053071	aroB	Dehydroquinate synthetase	
AF060558	hisH	Glutamine amidotransferase	
AF086704	hisE	Phosphoribosyl-ATP-pyrophosphohydrolase	
AF114233	aroA	5-enolpyruvylshikimate 3-phosphate synthase	
AF116184	panD	L-aspartate-alpha-decarboxylase precursor	Dusch, N. et al. "Expression of the Corynebacterium glutamicum panD gene encoding L-aspartate-alpha-decarboxylase leads to pantothenate overproduction in Escherichia coli," *Appl. Environ. Microbiol.*, 65(4)1530-1539 (1999)

GenBank™ Accession No.	Gene Name	Gene Function	Reference
AF124518	aroD; aroE	3-dehydroquinase; shikimate dehydrogenase	
AF124600	aroC; aroK; aroB; pepQ	Chorismate synthase; shikimate kinase; 3-dehydroquinate synthase; putative cytoplasmic peptidase	
AF145897	inhA		
AF145898	inhA		
AJ001436	ectP	Transport of ectoine, glycine betaine, proline	Peter, H. et al. "Corynebacterium glutamicum is equipped with four secondary carriers for compatible solutes. Identification, sequencing, and characterization of the proline/ectoine uptake system, ProP, and the ectoine/proline/glycine betaine carrier, EctP," *J. Bacteriol.*, 180(22):6005-6012 (1998)
AJ004934	dapD	Tetrahydrodipicolinate succinylase (incomplete')	Wehrmann, A. et al. "Different modes of diaminopimelate synthesis and their role in cell wall integrity: A study with Corynebacterium glutamicum," *J Bacteriol.*, 180(12):3159-3165 (1998)
AJ007732	ppc; secG; amt; ocd; soxA	Phosphoenolpyruvate-carboxylase; ?; high affinity ammonium uptake protein; putative ornithine-cyclodecarboxylase; sarcosine oxidase	
AJ010319	ftsY, glnB, glnD; srp; amtP	Involved in cell division; PII protein; uridylyltransferase (uridylyl-removing enzmye); signal recognition particle; low affinity ammonium uptake protein	Jakoby, M. et al. "Nitrogen regulation in Corynebacterium glutamicum; Isolation of genes involved in biochemical characterization of corresponding proteins," *FEMS Microbiol*, 173(2):303-310 (1999)
AJ132968	cat	Chloramphenicol aceteyl transferase	
AJ224946	mqo	L-malate: quinone oxidoreductase	Molenaar, D. et al. "Biochemical and genetic characterization of the membrane-associated malate dehydrogenase (acceptor) from Corynebacterium glutamicum," *Eur. J Biochem*, 254(2):395-403 (1998)
AJ238250	ndh	NADH dehydrogenase	
AJ238703	porA	Porin	Lichtinger, T. et al "Biochemical and biophysical characterization of the cell wall porin of Corynebacterium glutamicum: The channel is formed by a low molecular mass polypeptide," *Biochemistry*, 37(43):15024-15032 (1998)
D17429		Transposable element IS31831	Vertes, A.A. et al. "Isolation and characterization of IS31831, a transposable element from Corynebacterium glutamicum," *Mol Microbiol*, 11(4):739-746 (1994)

GenBank™ Accession No.	Gene Name	Gene Function	Reference
D84102	odhA	2-oxoglutarate dehydrogenase	Usuda, Y. et al. "Molecular cloning of the Corynebacterium glutamicum (Brevibacterium lactofermentum AJ12036) odhA gene encoding a novel type of 2-oxoglutarate dehydrogenase," *Microbiology*, 142.3347-3354 (1996)
E01358	hdh, hk	Homoserine dehydrogenase; homoserine kinase	Katsumata, R. et al. "Production of L-threonine and L-isoleucine," Patent: JP 1987232392-A 1 10/12/87
E01359		Upstream of the start codon of homoserine kinase gene	Katsumata, R. et al. "Production of L-threonine and L-isoleucine," Patent: JP 1987232392-A 2 10/12/87
E01375		Tryptophan operon	
E01376	trpL; trpE	Leader peptide; anthranilate synthase	Matsui, K. et al. "Tryptophan operon, peptide and protein coded thereby, utilization of tryptophan operon gene expression and production of tryptophan," Patent: JP 1987244382-A 1 10/24/87
E01377		Promoter and operator regions of tryptophan operon	Matsui, K. et al. "Tryptophan operon, peptide and protein coded thereby, utilization of tryptophan operon gene expression and production of tryptophan," Patent: JP 1987244382-A 1 10/24/87
E03937		Biotin-synthase	Hatakeyama, K. et al "DNA fragment containing gene capable of coding biotin synthetase and its utilization," Patent: JP 1992278088-A 1 10/02/92
E04040		Diamino pelargonic acid aminotransferase	Kohama, K. et al. "Gene coding diaminopelargonic acid aminotransferase and desthiobiotin synthetase and its utilization," Patent: JP 1992330284-A 1 11/18/92
E04041		Desthiobiotinsynthetase	Kohama, K. et al. "Gene coding diaminopelargonic acid aminotransferase and desthiobiotin synthetase and its utilization," Patent: JP 1992330284-A 1 11/18/92
E04307		Flavum aspartase	Kurusu, Y. et al "Gene DNA coding aspartase and utilization thereof," Patent: JP 1993030977-A 1 02/09/93
E04376		Isocitric acid lyase	Katsumata, R. et al. "Gene manifestation controlling DNA," Patent: JP 1993056782-A 3 03/09/93
E04377		Isocitric acid lyase N-terminal fragment	Katsumata, R et al. "Gene manifestation controlling DNA," Patent. JP 1993056782-A 3 03/09/93
E04484		Prephenate dehydratase	Sotouchi, N. et al "Production of L-phenylalanine by fermentation," Patent. JP 1993076352-A 2 03/30/93
E05108		Aspartokinase	Fugono, N et al. "Gene DNA coding Aspartokinase and its use," Patent: JP 1993184366-A 1 07/27/93
E05112		Dihydro-dipichorinate synthetase	Hatakeyama, K. et al "Gene DNA coding dihydrodipicolinic acid synthetase and its use," Patent: JP 1993184371-A 1 07/27/93

GenBank™ Accession No.	Gene Name	Gene Function	Reference
E05776		Diaminopimelic acid dehydrogenase	Kobayashi, M et al. "Gene DNA coding Diaminopimelic acid dehydrogenase and its use," Patent. JP 1993284970-A 1 11/02/93
E05779		Threonine synthase	Kohama, K. et al. "Gene DNA coding threonine synthase and its use," Patent. JP 1993284972-A 1 11/02/93
E06110		Prephenate dehydratase	Kikuchi, T. et al. "Production of L-phenylalanine by fermentation method," Patent. JP 1993344881-A 1 12/27/93
E06111		Mutated Prephenate dehydratase	Kikuchi, T. et al. "Production of L-phenylalanine by fermentation method," Patent: JP 1993344881-A 1 12/27/93
E06146		Acetohydroxy acid synthetase	Inui, M. et al. "Gene capable of coding Acetohydroxy acid synthetase and its use," Patent JP 1993344893-A 1 12/27/93
E06825		Aspartokinase	Sugimoto, M. et al "Mutant aspartokinase gene," patent: JP 1994062866-A 1 03/08/94
E06826		Mutated aspartokinase alpha subunit	Sugimoto, M et al. "Mutant aspartokinase gene," patent: JP 1994062866-A 1 03/08/94
E06827		Mutated aspartokinase alpha subunit	Sugimoto, M. et al. "Mutant aspartokinase gene," patent: JP 1994062866-A 1 03/08/94
E07701	secY		Honno, N. et al. "Gene DNA participating in integration of membraneous protein to membrane," Patent: JP 1994169780-A 1 06/21/94
E08177		Aspartokinase	Sato, Y et al "Genetic DNA capable of coding Aspartokinase released from feedback inhibition and its utilization," Patent: JP 1994261766-A 1 09/20/94
E08178, E08179, E08180, E08181, E08182		Feedback inhibition-released Aspartokinase	Sato, Y. et al. "Genetic DNA capable of coding Aspartokinase released from feedback inhibition and its utilization," Patent: JP 1994261766-A 1 09/20/94
E08232		Acetohydroxy-acid isomeroreductase	Inui, M. et al. "Gene DNA coding acetohydroxy acid isomeroreductase," Patent: JP 1994277067-A 1 10/04/94
E08234	secE		Asai, Y. et al. "Gene DNA coding for translocation machinery of protein," Patent: JP 1994277073-A 1 10/04/94
E08643		FT aminotransferase and desthiobiotin synthetase promoter region	Hatakeyama, K. et al. "DNA fragment having promoter function in coryneform bacterium," Patent: JP 1995031476-A 1 02/03/95
E08646		Biotin synthetase	Hatakeyama, K. et al. "DNA fragment having promoter function in coryneform bacterium," Patent: JP 1995031476-A 1 02/03/95

GenBank™ Accession No.	Gene Name	Gene Function	Reference
E08649		Aspartase	Kohama, K. et al "DNA fragment having promoter function in coryneform bacterium," Patent. JP 1995031478-A 1 02/03/95
E08900		Dihydrodipicolinate reductase	Madori, M. et al "DNA fragment containing gene coding Dihydrodipicolinate acid reductase and utilization thereof," Patent: JP 1995075578-A 1 03/20/95
E08901		Diaminopimelic acid decarboxylase	Madori, M. et al. "DNA fragment containing gene coding Diaminopimelic acid decarboxylase and utilization thereof," Patent. JP 1995075579-A 1 03/20/95
E12594		Serine hydroxymethyltransferase	Hatakeyama, K. et al. "Production of L-trypophan," Patent: JP 1997028391-A 1 02/04/97
E12760, E12759, E12758		transposase	Moriya, M. et al "Amplification of gene using artificial transposon," Patent: JP 1997070291-A 03/18/97
E12764		Arginyl-tRNA synthetase; diaminopimelic acid decarboxylase	Moriya, M. et al. "Amplification of gene using artificial transposon," Patent. JP 1997070291-A 03/18/97
E12767		Dihydrodipicolinic acid synthetase	Moriya, M. et al. "Amplification of gene using artificial transposon," Patent: JP 1997070291-A 03/18/97
E12770		aspartokinase	Moriya, M. et al. "Amplification of gene using artificial transposon," Patent: JP 1997070291-A 03/18/97
E12773		Dihydrodipicolinic acid reductase	Moriya, M. et al. "Amplification of gene using artificial transposon," Patent: JP 1997070291-A 03/18/97
E13655		Glucose-6-phosphate dehydrogenase	Hatakeyama, K et al. "Glucose-6-phosphate dehydrogenase and DNA capable of coding the same," Patent: JP 1997224661-A 1 09/02/97
L01508	IlvA	Threonine dehydratase	Moeckel, B. et al. "Functional and structural analysis of the threonine dehydratase of Corynebacterium glutamicum," *J. Bacteriol*, 174.8065-8072 (1992)
L07603	EC 4.2 1.15	3-deoxy-D-arabinoheptulosonate-7-phosphate synthase	Chen, C et al. "The cloning and nucleotide sequence of Corynebacterium glutamicum 3-deoxy-D-arabinoheptulosonate-7-phosphate synthase gene," *FEMS Microbiol Lett.*, 107:223-230 (1993)
L09232	IlvB; ilvN; ilvC	Acetohydroxy acid synthase large subunit; Acetohydroxy acid synthase small subunit; Acetohydroxy acid isomeroreductase	Keilhauer, C. et al. "Isoleucine synthesis in Corynebacterium glutamicum. molecular analysis of the ilvB-ilvN-ilvC operon," *J. Bacteriol.*, 175(17).5595-5603 (1993)

GenBank™ Accession No.	Gene Name	Gene Function	Reference
L18874	PtsM	Phosphoenolpyruvate sugar phosphotransferase	Fouet, A et al "Bacillus subtilis sucrose-specific enzyme II of the phosphotransferase system: expression in Escherichia coli and homology to enzymes II from enteric bacteria," *PNAS USA*, 84(24) 8773-8777 (1987); Lee, J.K. et al. "Nucleotide sequence of the gene encoding the Corynebacterium glutamicum mannose enzyme II and analyses of the deduced protein sequence," *FEMS Microbiol Lett*, 119(1-2).137-145 (1994)
L27123	aceB	Malate synthase	Lee, H-S. et al. "Molecular characterization of aceB, a gene encoding malate synthase in Corynebacterium glutamicum," *J Microbiol. Biotechnol*, 4(4) 256-263 (1994)
L27126		Pyruvate kinase	Jetten, M. S. et al. "Structural and functional analysis of pyruvate kinase from Corynebacterium glutamicum," *Appl. Environ Microbiol.*, 60(7):2501-2507 (1994)
L28760	aceA	Isocitrate lyase	
L35906	dtxr	Diphtheria toxin repressor	Oguiza, J.A. et al. "Molecular cloning, DNA sequence analysis, and characterization of the Corynebacterium diphtheriae dtxR from Brevibacterium lactofermentum," *J. Bacteriol*, 177(2):465-467 (1995)
M13774		Prephenate dehydratase	Follettie, M.T. et al. "Molecular cloning and nucleotide sequence of the Corynebacterium glutamicum pheA gene," *J Bacteriol.*, 167:695-702 (1986)
M16175	5S rRNA		Park, Y-H et al. "Phylogenetic analysis of the coryneform bacteria by 56 rRNA sequences," *J. Bacteriol*, 169:1801-1806 (1987)
M16663	trpE	Anthranilate synthase, 5' end	Sano, K. et al "Structure and function of the trp operon control regions of Brevibacterium lactofermentum, a glutamic-acid-producing bacterium," *Gene*, 52.191-200 (1987)
M16664	trpA	Tryptophan synthase, 3' end	Sano, K. et al. "Structure and function of the trp operon control regions of Brevibacterium lactofermentum, a glutamic-acid-producing bacterium," *Gene*, 52.191-200 (1987)
M25819		Phosphoenolpyruvate carboxylase	O'Regan, M. et al. "Cloning and nucleotide sequence of the Phosphoenolpyruvate carboxylase-coding gene of Corynebacterium glutamicum ATCC13032," *Gene*, 77(2):237-251 (1989)
M85106		23S rRNA gene insertion sequence	Roller, C. et al. "Gram-positive bacteria with a high DNA G+C content are characterized by a common insertion within their 23S rRNA genes," *J. Gen Microbiol*, 138.1167-1175 (1992)

GenBank™ Accession No.	Gene Name	Gene Function	Reference
M85107, M85108		23S rRNA gene insertion sequence	Roller, C. et al. "Gram-positive bacteria with a high DNA G+C content are characterized by a common insertion within their 23S rRNA genes," *J. Gen Microbiol*, 138:1167-1175 (1992)
M89931	aecD; bmQ, yhbw	Beta C-S lyase, branched-chain amino acid uptake carrier, hypothetical protein yhbw	Rossol, I. et al "The Corynebacterium glutamicum aecD gene encodes a C-S lyase with alpha, beta-elimination activity that degrades aminoethylcysteine," *J. Bacteriol.*, 174(9):2968-2977 (1992); Tauch, A. et al. "Isoleucine uptake in Corynebacterium glutamicum ATCC 13032 is directed by the bmQ gene product," *Arch Microbiol.*, 169(4):303-312 (1998)
S59299	trp	Leader gene (promoter)	Henry, D.M. et al. "Cloning of the trp gene cluster from a tryptophan-hyperproducing strain of Corynebacterium glutamicum: identification of a mutation in the trp leader sequence," *Appl. Environ. Microbiol*, 59(3):791-799 (1993)
U11545	trpD	Anthranilate phosphoribosyltransferase	O'Gara, J.P. and Dunican, L.K (1994) Complete nucleotide sequence of the Corynebacterium glutamicum ATCC 21850 tpD gene." Thesis, Microbiology Department, University College Galway, Ireland.
U13922	cglIM; cglIR, clglIR	Putative type II 5-cytosoine methyltransferase; putative type II restriction endonuclease; putative type I or type III restriction endonuclease	Schafer, A. et al "Cloning and characterization of a DNA region encoding a stress-sensitive restriction system from Corynebacterium glutamicum ATCC 13032 and analysis of its role in intergeneric conjugation with Escherichia coli," *J Bacteriol.*, 176(23):7309-7319 (1994); Schafer, A. et al. "The Corynebacterium glutamicum cglIM gene encoding a 5-cytosine in an McrBC-deficient Escherichia coli strain," *Gene*, 203(2):95-101 (1997)
U14965	recA		
U31224	ppx		Ankri, S et al. "Mutations in the Corynebacterium glutamicumproline biosynthetic pathway: A natural bypass of the proA step," *J Bacteriol.*, 178(15):4412-4419 (1996)
U31225	proC	L-proline: NADP+ 5-oxidoreductase	Ankri, S. et al "Mutations in the Corynebacterium glutamicumproline biosynthetic pathway. A natural bypass of the proA step," *J Bacteriol*, 178(15):4412-4419 (1996)
U31230	obg; proB, unkdh	?;gamma glutamyl kinase;similar to D-isomer specific 2-hydroxyacid dehydrogenases	Ankri, S. et al. "Mutations in the Corynebacterium glutamicumproline biosynthetic pathway. A natural bypass of the proA step," *J Bacteriol*, 178(15):4412-4419 (1996)

GenBank™ Accession No.	Gene Name	Gene Function	Reference
U31281	bioB	Biotin synthase	Serebriiskii, I.G., "Two new members of the bio B superfamily: Cloning, sequencing and expression of bio B genes of Methylobacillus flagellatum and Corynebacterium glutamicum," *Gene*, 175:15-22 (1996)
U35023	thtR; accBC	Thiosulfate sulfurtransferase; acyl CoA carboxylase	Jager, W. et al. "A Corynebacterium glutamicum gene encoding a two-domain protein similar to biotin carboxylases and biotin-carboxyl-carrier proteins," *Arch Microbiol.*, 166(2);76-82 (1996)
U43535	cmr	Multidrug resistance protein	Jager, W. et al. "A Corynebacterium glutamicum gene conferring multidrug resistance in the heterologous host Escherichia coli," *J Bacteriol.*, 179(7):2449-2451 (1997)
U43536	clpB	Heat shock ATP-binding protein	
U53587	aphA-3	3'5''-aminoglycoside phosphotransferase	
U89648		Corynebacterium glutamicum unidentified sequence involved in histidine biosynthesis, partial sequence	
X04960	trpA; trpB; trpC; trpD; trpE; trpG; trpL	Tryptophan operon	Matsui, K. et al. "Complete nucleotide and deduced amino acid sequences of the Brevibacterium lactofermentum tryptophan operon," *Nucleic Acids Res.*, 14(24):10113-10114 (1986)
X07563	lys A	DAP decarboxylase (meso-diaminopimelate decarboxylase, EC 4.1.1.20)	Yeh, P. et al. "Nucleic sequence of the lysA gene of Corynebacterium glutamicum and possible mechanisms for modulation of its expression," *Mol Gen. Genet.*, 212(1):112-119 (1988)
X14234	EC 4 1.1.31	Phosphoenolpyruvate carboxylase	Eikmanns, B.J. et al. "The Phosphoenolpyruvate carboxylase gene of Corynebacterium glutamicum: Molecular cloning, nucleotide sequence, and expression," *Mol Gen. Genet.*, 218(2):330-339 (1989); Lepiniec, L. et al. "Sorghum Phosphoenolpyruvate carboxylase gene family: structure, function and molecular evolution," *Plant Mol. Biol.*, 21 (3):487-502 (1993)
X17313	fda	Fructose-bisphosphate aldolase	Von der Osten, C.H. et al. "Molecular cloning, nucleotide sequence and fine-structural analysis of the Corynebacterium glutamicum fda gene: structural comparison of C. glutamicum fructose-1, 6-biphosphate aldolase to class I and class II aldolases," *Mol. Microbiol.*,
X53993	dapA	L-2, 3-dihydrodipicolinate synthetase (EC 4 2.1.52)	Bonnassie, S. et al. "Nucleic sequence of the dapA gene from Corynebacterium glutamicum," *Nucleic Acids Res.*, 18(21):6421 (1990)

GenBank™ Accession No.	Gene Name	Gene Function	Reference
X54223		AttB-related site	Cianciotto, N. et al. "DNA sequence homology between att B-related sites of Corynebacterium diphtheriae, Corynebacterium ulcerans, Corynebacterium glutamicum , and the attP site of lambdacorynephage," *FEMS Microbiol. Lett* , 66:299-302 (1990)
X54740	argS; lysA	Arginyl-tRNA synthetase; Diaminopimelate decarboxylase	Marcel, T. et al. "Nucleotide sequence and organization of the upstream region of the Corynebacterium glutamicum lysA gene," *Mol Microbiol* , 4(11):1819-1830 (1990)
X55994	trpL; trpE	Putative leader peptide; anthranilate synthase component I	Heery, D.M. et al. "Nucleotide sequence of the Corynebacterium glutamicum trpE gene," *Nucleic Acids Res.*, 18(23):7138 (1990)
X56037	thrC	Threonine synthase	Han, K.S. et al. "The molecular structure of the Corynebacterium glutamicum threonine synthase gene," *Mol Microbiol* , 4(10).1693-1702 (1990)
X56075	attB-related site	Attachment site	Cianciotto, N. et al. "DNA sequence homology between att B-related sites of Corynebacterium diphtheriae, Corynebacterium ulcerans, Corynebacterium glutamicum , and the attP site of lambdacorynephage," *FEMS Microbiol. Lett.*, 66:299-302 (1990)
X57226	lysC-alpha; lysC-beta; asd	Aspartokinase-alpha subunit, Aspartokinase-beta subunit; aspartate beta semialdehyde dehydrogenase	Kalinowski, J. et al. "Genetic and biochemical analysis of the Aspartokinase from Corynebacterium glutamicum," *Mol Microbiol.*, 5(5).1197-1204 (1991); Kalinowski, J. et al. "Aspartokinase genes lysC alpha and lysC beta overlap and are adjacent to the aspertate beta-semialdehyde dehydrogenase gene asd in Corynebacterium glutamicum," *Mol Gen Genet* , 224(3):317-324 (1990)
X59403	gap,pgk; tpi	Glyceraldehyde-3-phosphate; phosphoglycerate kinase, triosephosphate isomerase	Eikmanns, B.J. "Identification, sequence analysis, and expression of a Corynebacterium glutamicum gene cluster encoding the three glycolytic enzymes glyceraldehyde-3-phosphate dehydrogenase, 3-phosphoglycerate kinase, and triosephosphate isomeras," *J. Bacteriol* , 174(19):6076-6086 (1992)
X59404	gdh	Glutamate dehydrogenase	Bormann, E.R. et al. "Molecular analysis of the Corynebacterium glutamicum gdh gene encoding glutamate dehydrogenase," *Mol Microbiol.*, 6(3):317-326 (1992)
X60312	lysI	L-lysine permease	Seep-Feldhaus, A.H. et al. "Molecular analysis of the Corynebacterium glutamicum lysI gene involved in lysine uptake," *Mol Microbiol* , 5(12).2995-3005 (1991)

GenBank™ Accession No.	Gene Name	Gene Function	Reference
X66078	cop1	Ps1 protein	Joliff, G. et al "Cloning and nucleotide sequence of the csp1 gene encoding PS1, one of the two major secreted proteins of Corynebacterium glutamicum: The deduced N-terminal region of PS1 is similar to the Mycobacterium antigen 85 complex," *Mol. Microbiol*, 6(16).2349-2362 (1992)
X66112	gli	Citrate synthase	Eikmanns, B.J. et al. "Cloning sequence, expression and transcriptional analysis of the Corynebacterium glutamicum gltA gene encoding citrate synthase," *Microbiol*, 140:1817-1828 (1994)
X67737	dapB	Dihydrodipicolinate reductase	
X69103	csp2	Surface layer protein PS2	Peyret, J.L. et al. "Characterization of the cspB gene encoding PS2, an ordered surface-layer protein in Corynebacterium glutamicum," *Mol Microbiol*, 9(1):97-109 (1993)
X69104		IS3 related insertion element	Bonamy, C et al. "Identification of IS1206, a Corynebacterium glutamicum IS3-related insertion sequence and phylogenetic analysis," *Mol. Microbiol*, 14(3):571-581 (1994)
X70959	leuA	Isopropylmalate synthase	Patek, M. et al. "Leucine synthesis in Corynebacterium glutamicum: enzyme activities, structure of leuA, and effect of leuA inactivation on lysine synthesis," *Appl Environ. Microbiol*, 60(1).133-140 (1994)
X71489	icd	Isocitrate dehydrogenase (NADP+)	Eikmanns, B.J. et al "Cloning sequence analysis, expression, and inactivation of the Corynebacterium glutamicum icd gene encoding isocitrate dehydrogenase and biochemical characterization of the enzyme," *J. Bacteriol*, 177(3).774-782 (1995)
X72855	GDHA	Glutamate dehydrogenase (NADP+)	
X75083, X70584	mtrA	5-methyltryptophan resistance	Heery, D.M. et al. "A sequence from a tryptophan-hyperproducing strain of Corynebacterium glutamicum encoding resistance to 5-methyltryptophan," *Biochem Biophys Res. Commun*, 201(3):1255-1262 (1994)
X75085	recA		Fitzpatrick, R. et al. "Construction and characterization of recA mutant strains of Corynebacterium glutamicum and Brevibacterium lactofermentum," *Appl. Microbiol Biotechnol*, 42(4).575-580 (1994)
X75504	aceA; thiX	Partial Isocitrate lyase; ?	Reinscheid, D.J. et al. "Characterization of the isocitrate lyase gene from Corynebacterium glutamicum and biochemical analysis of the enzyme," *J. Bacteriol*, 176(12):3474-3483 (1994)
X76875		ATPase beta-subunit	Ludwig, W et al. "Phylogenetic relationships of bacteria based on comparative sequence analysis of elongation factor Tu and ATP-synthase beta-subunit genes," *Antonie Van Leeuwenhoek*, 64:285-305 (1993)

GenBank™ Accession No.	Gene Name	Gene Function	Reference
X77034	tuf	Elongation factor Tu	Ludwig, W. et al. "Phylogenetic relationships of bacteria based on comparative sequence analysis of elongation factor Tu and ATP-synthase beta-subunit genes," *Antonie Van Leeuwenhoek*, 64:285-305 (1993)
X77384	recA		Billman-Jacobe, H. "Nucleotide sequence of a recA gene from Corynebacterium glutamicum," *DNA Seq.*, 4(6):403-404 (1994)
X78491	aceB	Malate synthase	Reinscheid, D.J. et al. "Malate synthase from Corynebacterium glutamicum pta-ack operon encoding phosphotransacetylase: sequence analysis," *Microbiology*, 140:3099-3108 (1994)
X80629	16S rDNA	16S ribosomal RNA	Rainey, F.A. et al. "Phylogenetic analysis of the genera Rhodococcus and Norcardia and evidence for the evolutionary origin of the genus Norcardia from within the radiation of Rhodococcus species," *Microbiol.*, 141:523-528 (1995)
X81191	gluA; gluB; gluC; gluD	Glutamate uptake system	Kronemeyer, W. et al "Structure of the gluABCD cluster encoding the glutamate uptake system of Corynebacterium glutamicum," *J Bacteriol*, 177(5):1152-1158 (1995)
X81379	dapE	Succinyldiaminopimelate desuccinylase	Wehrmann, A et al "Analysis of different DNA fragments of Corynebacterium glutamicum complementing dapE of Escherichia coli," *Microbiology*, 40:3349-56 (1994)
X82061	16S rDNA	16S ribosomal RNA	Ruimy, R et al. "Phylogeny of the genus Corynebacterium deduced from analyses of small-subunit ribosomal DNA sequences," *Int. J Syst Bacteriol*, 45(4):740-746 (1995)
X82928	asd; lysC	Aspartate-semialdehyde dehydrogenase; ?	Serebrijski, I. et al. "Multicopy suppression by asd gene and osmotic stress-dependent complementation by heterologous proA in proA mutants," *J. Bacteriol*, 177(24):7255-7260 (1995)
X82929	proA	Gamma-glutamyl phosphate reductase	Serebrijski, I. et al. "Multicopy suppression by asd gene and osmotic stress-dependent complementation by heterologous proA in proA mutants," *J Bacteriol*, 177(24):7255-7260 (1995)
X84257	16S rDNA	16S ribosomal RNA	Pascual, C. et al. "Phylogenetic analysis of the genus Corynebacterium based on 16S rRNA gene sequences," *Int. J Syst Bacteriol.*, 45(4):724-728 (1995)
X85965	aroP; dapE	Aromatic amino acid permease; ?	Wehrmann, A. et al. "Functional analysis of sequences adjacent to dapE of Corynebacterium glutamicumproline reveals the presence of aroP, which encodes the aromatic amino acid transporter," *J. Bacteriol*, 177(20):5991-5993 (1995)

GenBank™ Accession No.	Gene Name	Gene Function	Reference
X86157	argB, argC; argD; argF; argJ	Acetylglutamate kinase; N-acetyl-gamma-glutamyl-phosphate reductase; acetylornithine aminotransferase; ornithine carbamoyltransferase; glutamate N-acetyltransferase	Sakanyan, V. et al. "Genes and enzymes of the acetyl cycle of arginine biosynthesis in Corynebacterium glutamicum: enzyme evolution in the early steps of the arginine pathway," *Microbiology*, 142:99-108 (1996)
X89084	pta; ackA	Phosphate acetyltransferase, acetate kinase	Reinscheid, D.J. et al. "Cloning, sequence analysis, expression and inactivation of the Corynebacterium glutamicum pta-ack operon encoding phosphotransacetylase and acetate kinase," *Microbiology*, 145:503-513 (1999)
X89850	attB	Attachment site	Le Marrec, C. et al. "Genetic characterization of site-specific integration functions of phi AAU2 infecting "Arthrobacter aureus C70," *J. Bacteriol.*, 178(7):1996-2004 (1996)
X90356		Promoter fragment F1	Patek, M. et al. "Promoters from Corynebacterium glutamicum: cloning, molecular analysis and search for a consensus motif," *Microbiology*, 142:1297-1309 (1996)
X90357		Promoter fragment F2	Patek, M. et al. "Promoters from Corynebacterium glutamicum: cloning, molecular analysis and search for a consensus motif," *Microbiology*, 142:1297-1309 (1996)
X90358		Promoter fragment F10	Patek, M. et al. "Promoters from Corynebacterium glutamicum: cloning, molecular analysis and search for a consensus motif," *Microbiology*, 142:1297-1309 (1996)
X90359		Promoter fragment F13	Patek, M. et al. "Promoters from Corynebacterium glutamicum: cloning, molecular analysis and search for a consensus motif," *Microbiology*, 142:1297-1309 (1996)
X90360		Promoter fragment F22	Patek, M. et al. "Promoters from Corynebacterium glutamicum: cloning, molecular analysis and search for a consensus motif," *Microbiology*, 142:1297-1309 (1996)
X90361		Promoter fragment F34	Patek, M. et al. "Promoters from Corynebacterium glutamicum: cloning, molecular analysis and search for a consensus motif," *Microbiology*, 142:1297-1309 (1996)
X90362		Promoter fragment F37	Patek, M. et al. "Promoters from Corynebacterium glutamicum: cloning, molecular analysis and search for a consensus motif," *Microbiology*, 142:1297-1309 (1996)

GenBank™ Accession No.	Gene Name	Gene Function	Reference
X90363		Promoter fragment F45	Patek, M et al "Promoters from Corynebacterium glutamicum: cloning, molecular analysis and search for a consensus motif," *Microbiology*, 142:1297-1309 (1996)
X90364		Promoter fragment F64	Patek, M. et al. "Promoters from Corynebacterium glutamicum: cloning, molecular analysis and search for a consensus motif," *Microbiology*, 142:1297-1309 (1996)
X90365		Promoter fragment F75	Patek, M. et al. "Promoters from Corynebacterium glutamicum: cloning, molecular analysis and search for a consensus motif," *Microbiology*, 142:1297-1309 (1996)
X90366		Promoter fragment PF101	Patek, M. et al. "Promoters from Corynebacterium glutamicum: cloning, molecular analysis and search for a consensus motif," *Microbiology*, 142:1297-1309 (1996)
X90367		Promoter fragment PF104	Patek, M. et al. "Promoters from Corynebacterium glutamicum: cloning, molecular analysis and search for a consensus motif," *Microbiology*, 142:1297-1309 (1996)
X90368		Promoter fragment PF109	Patek, M. et al. "Promoters from Corynebacterium glutamicum: cloning, molecular analysis and search for a consensus motif," *Microbiology*, 142:1297-1309 (1996)
X93513	amt	Ammonium transport system	Siewe, R M. et al. "Functional and genetic characterization of the (methyl) ammonium uptake carrier of Corynebacterium glutamicum," *J. Biol Chem*, 271(10):5398-5403 (1996)
X93514	betP	Glycine betaine transport system	Peter, H et al. "Isolation, characterization, and expression of the Corynebacterium glutamicum betP gene, encoding the transport system for the compatible solute glycine betaine," *J Bacteriol*, 178(17):5229-5234 (1996)
X95649	orf4		Patek, M. et al. "Identification and transcriptional analysis of the dapB-ORF2-dapA-ORF4 operon of Corynebacterium glutamicum, encoding two enzymes involved in L-lysine synthesis," *Biotechnol Lett*, 19:1113-1117 (1997)
X96471	lysE; lysG	Lysine exporter protein, Lysine export regulator protein	Vrljic, M. et al. "A new type of transporter with a new type of cellular function: L-lysine export from Corynebacterium glutamicum," *Mol Microbiol*, 22(5):815-826 (1996)

GenBank™ Accession No.	Gene Name	Gene Function	Reference
X96580	panB, panC; xylB	3-methyl-2-oxobutanoate hydroxymethyltransferase; pantoate-beta-alanine ligase; xylulokinase	Sahm, H et al. "D-pantothenate synthesis in Corynebacterium glutamicum and use of panBC and genes encoding L-valine synthesis for D-pantothenate overproduction," *Appl Environ Microbiol*, 65(5).1973-1979 (1999)
X96962		Insertion sequence IS1207 and transposase	
X99289		Elongation factor P	Ramos, A. et al. "Cloning, sequencing and expression of the gene encoding elongation factor P in the amino-acid producer Brevibacterium lactofermentum (Corynebacterium glutamicum ATCC 13869)," *Gene*, 198:217-222 (1997)
Y00140	thrB	Homoserine kinase	Mateos, L.M. et al. "Nucleotide sequence of the homoserine kinase (thrB) gene of the Brevibacterium lactofermentum," *Nucleic Acids Res.*, 15(9):3922 (1987)
Y00151	ddh	Meso-diaminopimelate D-dehydrogenase (EC 1.4.1.16)	Ishino, S. et al. "Nucleotide sequence of the meso-diaminopimelate D-dehydrogenase gene from Corynebacterium glutamicum," *Nucleic Acids Res.*, 15(9):3917 (1987)
Y00476	thrA	Homoserine dehydrogenase	Mateos, L.M. et al "Nucleotide sequence of the homoserine dehydrogenase (thrA) gene of the Brevibacterium lactofermentum," *Nucleic Acids Res.*, 15(24):10598 (1987)
Y00546	hom; thrB	Homoserine dehydrogenase; homoserine kinase	Peoples, O.P. et al. "Nucleotide sequence and fine structural analysis of the Corynebacterium glutamicum hom-thrB operon," *Mol Microbiol.*, 2(1):63-72 (1988)
Y08964	murC, ftsQ/divD; ftsZ	UPD-N-acetylmuramate-alanine ligase, division initiation protein or cell division protein; cell division protein	Honrubia, M P. et al. "Identification, characterization, and chromosomal organization of the ftsZ gene from Brevibacterium lactofermentum," *Mol Gen Genet*, 259(1):97-104 (1998)
Y09163	putP	High affinity proline transport system	Peter, H. et al. "Isolation of the putP gene of Corynebacterium glutamicumproline and characterization of a low-affinity uptake system for compatible solutes," *Arch Microbiol.*, 168(2):143-151 (1997)
Y09548	pyc	Pyruvate carboxylase	Peters-Wendisch, P.G. et al. "Pyruvate carboxylase from Corynebacterium glutamicum: characterization, expression and inactivation of the pyc gene," *Microbiology*, 144:915-927 (1998)
Y09578	leuB	3-isopropylmalate dehydrogenase	Patek, M. et al. "Analysis of the leuB gene from Corynebacterium glutamicum," *Appl Microbiol. Biotechnol.*, 50(1):42-47 (1998)
Y12472		Attachment site bacteriophage Phi-16	Moreau, S. et al. "Site-specific integration of corynephage Phi-16: The construction of an integration vector," *Microbiol.*, 145:539-548 (1999)

GenBank™ Accession No.	Gene Name	Gene Function	Reference
Y12537	proP	Proline/ectoine uptake system protein	Peter, H. et al. "Corynebacterium glutamicum is equipped with four secondary carriers for compatible solutes. Identification, sequencing, and characterization of the proline/ectoine uptake system, ProP, and the ectoine/proline/glycine betaine carrier, EctP," *J. Bacteriol.*, 180(22):6005-6012 (1998)
Y13221	glnA	Glutamine synthetase I	Jakoby, M. et al. "Isolation of Corynebacterium glutamicum glnA gene encoding glutamine synthetase I," *FEMS Microbiol Lett.*, 154(1):81-88 (1997)
Y16642	lpd	Dihydrolipoamide dehydrogenase	
Y18059		Attachment site Corynephage 304L	Moreau, S. et al "Analysis of the integration functions of φ304L. An integrase module among corynephages," *Virology*, 255(1) 150-159 (1999)
Z21501	argS; lysA	Arginyl-tRNA synthetase; diaminopimelate decarboxylase (partial)	Oguiza, J A et al. "A gene encoding arginyl-tRNA synthetase is located in the upstream region of the lysA gene in Brevibacterium lactofermentum. Regulation of argS-lysA cluster expression by arginine," *J Bacteriol.*,175(22):7356-7362 (1993)
Z21502	dapA; dapB	Dihydrodipicolinate synthase; dihydrodipicolinate reductase	Pisabarro, A et al. "A cluster of three genes (dapA, orf2, and dapB) of Brevibacterium lactofermentum encodes dihydrodipicolinate reductase, and a third polypeptide of unknown function," *J. Bacteriol.*, 175(9):2743-2749 (1993)
Z29563	thrC	Threonine synthase	Malumbres, M et al. "Analysis and expression of the thrC gene of the encoded threonine synthase," *Appl Environ Microbiol*, 60(7)2209-2219 (1994)
Z46753	16S rDNA	Gene for 16S ribosomal RNA	
Z49822	sigA	SigA sigma factor	Oguiza, J A. et al "Multiple sigma factor genes in Brevibacterium lactofermentum: Characterization of sigA and sigB," *J Bacteriol*, 178(2).550-553 (1996)
Z49823	galE; dtxR	Catalytic activity UDP-galactose 4-epimerase; diphtheria toxin regulatory protein	Oguiza, J A et al "The galE gene encoding the UDP-galactose 4-epimerase of Brevibacterium lactofermentum is coupled transcriptionally to the dtxR gene," *Gene*, 177.103-107 (1996)
Z49824	orf1; sigB	?; SigB sigma factor	Oguiza, J A. et al "Multiple sigma factor genes in Brevibacterium lactofermentum: Characterization of sigA and sigB," *J. Bacteriol*, 178(2):550-553 (1996)
Z66534		Transposase	Correia, A. et al. "Cloning and characterization of an IS-like element present in the genome of Brevibacterium lactofermentum ATCC 13869," *Gene*, 170(1) 91-94 (1996)

[1] A sequence for this gene was published in the indicated reference However, the sequence obtained by the inventors of the present application is significantly longer than the published version. It is believed that the published version relied on an incorrect start codon, and thus represents only a fragment of the actual coding region.

TABLE 3: Corynebacterium and Brevibacterium Strains Which May be Used in the Practice of the Invention

Genus	Species	ATCC	FERM	NRRL	CBCT	NCIMB	CBS	NCTC	DSMZ
Brevibacterium	ammoniagenes	21054							
Brevibacterium	ammoniagenes	19350							
Brevibacterium	ammoniagenes	19351							
Brevibacterium	ammoniagenes	19352							
Brevibacterium	ammoniagenes	19353							
Brevibacterium	ammoniagenes	19354							
Brevibacterium	ammoniagenes	19355							
Brevibacterium	ammoniagenes	19356							
Brevibacterium	ammoniagenes	21055							
Brevibacterium	ammoniagenes	21077							
Brevibacterium	ammoniagenes	21553							
Brevibacterium	ammoniagenes	21580							
Brevibacterium	ammoniagenes	39101							
Brevibacterium	butanicum	21196							
Brevibacterium	divaricatum	21792	P928						
Brevibacterium	flavum	21474							
Brevibacterium	flavum	21129							
Brevibacterium	flavum	21518							
Brevibacterium	flavum			B11474					
Brevibacterium	flavum			B11472					
Brevibacterium	flavum	21127							
Brevibacterium	flavum	21128							
Brevibacterium	flavum	21427							
Brevibacterium	flavum	21475							
Brevibacterium	flavum	21517							
Brevibacterium	flavum	21528							
Brevibacterium	flavum	21529							
Brevibacterium	flavum			B11477					

Brevibacterium	flavum			B11478					
Brevibacterium	flavum	21127							
Brevibacterium	flavum			B11474					
Brevibacterium	healii	15527							
Brevibacterium	ketoglutamicum	21004							
Brevibacterium	ketoglutamicum	21089							
Brevibacterium	ketosoreductum	21914							
Brevibacterium	lactofermentum				70				
Brevibacterium	lactofermentum				74				
Brevibacterium	lactofermentum				77				
Brevibacterium	lactofermentum	21798							
Brevibacterium	lactofermentum	21799							
Brevibacterium	lactofermentum	21800							
Brevibacterium	lactofermentum	21801							
Brevibacterium	lactofermentum			B11470					
Brevibacterium	lactofermentum			B11471					
Brevibacterium	lactofermentum	21086							
Brevibacterium	lactofermentum	21420							
Brevibacterium	lactofermentum	21086							
Brevibacterium	lactofermentum	31269							
Brevibacterium	linens	9174							
Brevibacterium	linens	19391							
Brevibacterium	linens	8377							
Brevibacterium	paraffinolyticum					11160			
Brevibacterium	spec.						717.73		
Brevibacterium	spec.						717.73		
Brevibacterium	spec.	14604							
Brevibacterium	spec.	21860							
Brevibacterium	spec.	21864							
Brevibacterium	spec.	21865							
Brevibacterium	spec.	21866							
Brevibacterium	spec.	19240							

Corynebacterium	acetoacidophilum	21476							
Corynebacterium	acetoacidophilum	13870							
Corynebacterium	acetoglutamicum			B11473					
Corynebacterium	acetoglutamicum			B11475					
Corynebacterium	acetoglutamicum	15806							
Corynebacterium	acetoglutamicum	21491							
Corynebacterium	acetoglutamicum	31270							
Corynebacterium	acetophilum			B3671					
Corynebacterium	ammoniagenes	6872						2399	
Corynebacterium	ammoniagenes	15511							
Corynebacterium	fujiokense	21496							
Corynebacterium	glutamicum	14067							
Corynebacterium	glutamicum	39137							
Corynebacterium	glutamicum	21254							
Corynebacterium	glutamicum	21255							
Corynebacterium	glutamicum	31830							
Corynebacterium	glutamicum	13032							
Corynebacterium	glutamicum	14305							
Corynebacterium	glutamicum	15455							
Corynebacterium	glutamicum	13058							
Corynebacterium	glutamicum	13059							
Corynebacterium	glutamicum	13060							
Corynebacterium	glutamicum	21492							
Corynebacterium	glutamicum	21513							
Corynebacterium	glutamicum	21526							
Corynebacterium	glutamicum	21543							
Corynebacterium	glutamicum	13287							
Corynebacterium	glutamicum	21851							
Corynebacterium	glutamicum	21253							
Corynebacterium	glutamicum	21514							
Corynebacterium	glutamicum	21516							
Corynebacterium	glutamicum	21299							

Corynebacterium	glutamicum	21300							
Corynebacterium	glutamicum	39684							
Corynebacterium	glutamicum	21488							
Corynebacterium	glutamicum	21649							
Corynebacterium	glutamicum	21650							
Corynebacterium	glutamicum	19223							
Corynebacterium	glutamicum	13869							
Corynebacterium	glutamicum	21157							
Corynebacterium	glutamicum	21158							
Corynebacterium	glutamicum	21159							
Corynebacterium	glutamicum	21355							
Corynebacterium	glutamicum	31808							
Corynebacterium	glutamicum	21674							
Corynebacterium	glutamicum	21562							
Corynebacterium	glutamicum	21563							
Corynebacterium	glutamicum	21564							
Corynebacterium	glutamicum	21565							
Corynebacterium	glutamicum	21566							
Corynebacterium	glutamicum	21567							
Corynebacterium	glutamicum	21568							
Corynebacterium	glutamicum	21569							
Corynebacterium	glutamicum	21570							
Corynebacterium	glutamicum	21571							
Corynebacterium	glutamicum	21572							
Corynebacterium	glutamicum	21573							
Corynebacterium	glutamicum	21579							
Corynebacterium	glutamicum	19049							
Corynebacterium	glutamicum	19050							
Corynebacterium	glutamicum	19051							
Corynebacterium	glutamicum	19052							
Corynebacterium	glutamicum	19053							
Corynebacterium	glutamicum	19054							

Corynebacterium	glutamicum	19055							
Corynebacterium	glutamicum	19056							
Corynebacterium	glutamicum	19057							
Corynebacterium	glutamicum	19058							
Corynebacterium	glutamicum	19059							
Corynebacterium	glutamicum	19060							
Corynebacterium	glutamicum	19185							
Corynebacterium	glutamicum	13286							
Corynebacterium	glutamicum	21515							
Corynebacterium	glutamicum	21527							
Corynebacterium	glutamicum	21544							
Corynebacterium	glutamicum	21492							
Corynebacterium	glutamicum			B8183					
Corynebacterium	glutamicum			B8182					
Corynebacterium	glutamicum			B12416					
Corynebacterium	glutamicum			B12417					
Corynebacterium	glutamicum			B12418					
Corynebacterium	glutamicum			B11476					
Corynebacterium	glutamicum	21608							
Corynebacterium	lilium		P973						
Corynebacterium	nitrilophilus	21419				11594			
Corynebacterium	spec.		P4445						
Corynebacterium	spec.		P4446						
Corynebacterium	spec.	31088							
Corynebacterium	spec.	31089							
Corynebacterium	spec.	31090							
Corynebacterium	spec.	31090							
Corynebacterium	spec	31090							
Corynebacterium	spec.	15954							20145
Corynebacterium	spec.	21857							
Corynebacterium	spec.	21862							
Corynebacterium	spec.	21863							

ATCC: American Type Culture Collection, Rockville, MD, USA

FERM: Fermentation Research Institute, Chiba, Japan

NRRL: ARS Culture Collection, Northern Regional Research Laboratory, Peoria, IL, USA

CECT: Coleccion Espanola de Cultivos Tipo, Valencia, Spain

NCIMB: National Collection of Industrial and Marine Bacteria Ltd., Aberdeen, UK

CBS: Centraalbureau voor Schimmelcultures, Baarn, NL

NCTC: National Collection of Type Cultures, London, UK

DSMZ: Deutsche Sammlung von Mikroorganismen und Zellkulturen, Braunschweig, Germany

For reference see Sugawara, H et al. (1993) World directory of collections of cultures of microorganisms: Bacteria, fungi and yeasts (4th edn), World federation for culture collections world data center on microorganisms, Saimata, Japan.

>>RXA00459-amino acid sequence
(1-987, translated) 329 residues
VCTR AAGGGAVTLK RARELRKKRG RMAARIADSV MAGELLHATG AIDRELNAVT RDSDRVVIAA
VRRSWATGFS RALMAMAASL GTVSIVISGH LEVSEVAGIM MLLGVLATPV AELGRVVEYR QNYKAATRIL
IPLLQRGSEF KHSQQKLPGL QATEGIPGVY VKGISALPGE RIYLHGSADA TRKWVTSLSA MEEGTDVIVN
GQRLSQLPLK QRRALIGIAS AHHHLSRGSV SRLVGLRVPD ATVEEIEQAL EQVGLNNTGK QRLKNGGHPW
STSQINKLKI ASATLRTPPL LVLEGITPEN LLNYPGVIIS TVQENPSETW RQVNI
>RXA00459-nucleotide sequence A: upstream
AGGCGTGGCTGTTACTGTGCCACTGCTCAT
>RXA00459-nucleotide sequence B: coding region
GTGTGTACCCGTGCTGCCGGTGGTGGCGCGGTGACTTTGAAAAGAGCACGTGAACTACGCAAAAAACGTGGACGCAT
GGCTGCGCGGATCGCAGATTCTGTCATGGCTGGAGAATTACTGCACGCAACAGGAGCAATAGACCGTGAGCTCAATG
CAGTCACCCGAGATTCCGACCGAGTGGTGATAGCTGCTGTAAGACGTTCCTGGGCCACCGGTTTTAGCCGCGCATTG
ATGGCCATGGCAGCCTCGCTTGGCACTGTCAGCATTGTGATTTCTGGCCACCTGGAAGTAAGTGAGGTTGCGGGAAT
AATGATGCTTCTTGGCGTTCTTGCCACTCCAGTTGCAGAACTTGGCCGCGTGGTGGAATATCGCCAAAATTATAAAG
CCGCGACACGCATCCTGATTCCACTTCTGCAACGAGGCTCAGAATTTAAACACTCCCAACAAAAACTACCCGGGTTG
CAAGCAACAGAAGGAATCCCCGGTGTCTATGTCAAAGGTATTTCCGCCCTTCCTGGAGAACGGATCTACCTCCACGG
CTCTGCAGATGCGACGAGAAAATGGGTCACCTCGTTGTCTGCAATGGAGGAAGGCACAGATGTAATAGTCAACGGTC
AAAGGCTTTCGCAGCTTCCTTTGAAACAACGACGCGCCCTCATCGGAATCGCCTCAGCACACCACCACTTAAGCCGT
GGTTCAGTATCGCGCCTGGTTGGTTTGCGAGTGCCGGATGCCACCGTGGAAGAAATTGAGCAAGCACTGGAACAAGT
TGGTCTGAACAACACCGGGAAACAACGCTTGAAAAACGGCGGACACCCCTGGAGTACTTCGCAGATCAACAAACTGA
AAATTGCCAGCGCCACCCTTCGAACCCCACCGCTTTTGGTACTTGAAGGCATCACCCCTGAAAACCTCCTCAACTAT
CCCGGAGTGATCATCTCCACCGTTCAGGAGAACCCATCCGAAACATGGCGGCAAGTGAACATC
>RXA00459-nucleotide sequence C: downstream
TAATCTAGAAACATGGCAGGACG

>>RXA00410-amino acid sequence
(1-666, translated) 222 residues
MMIYGKGSTE VRALDGISVQ IQSDKWTSIM GQSGSGKTTL LQCLSGLAQP TSGRVTLNKN NITLSSLSEN
KRAKLRRTHI SMVFQDFNLV PILSVKDNIL LPLRLAHRRV DKQWFEHITS VLKIDNRMRH LPGELSGGQQ
QRAAIARALM SRPDIVIADE PTGSLDSVTS DAVLNLFRSI VDDFGQSLVF VTHDKDAAHR GDVLITMRDG
KIIDTADLRV GR
>RXA00410-nucleotide sequence A: upstream
GTGTTGATGCGTTAGTCCACCCACGCAGCTACGCCCCAAAGGAATAATCTTGAACCCTGCCACAGATAACGCTCCGC
CGGTCCTTTCAGCCCAAGATCTC
>RXA00410-nucleotide sequence B: coding region
ATGATGATCTATGGAAAAGGATCAACAGAAGTTCGGGCTCTCGATGGCATTTCTGTACAGATTCAGTCCGACAAATG
GACCTCCATCATGGGGCAATCAGGCTCTGGCAAAACAACTCTGTTGCAGTGCCTTTCCGGATTGGCGCAGCCAACCT
CAGGCAGAGTGACACTGAACAAAAACAACATCACGTTGAGCTCCCTGTCAGAAAATAAGCGTGCCAAGCTGCGTCGC
ACGCACATCAGCATGGTGTTTCAGGATTTCAACTTGGTGCCTATTTTGTCGGTGAAGGACAATATTTTGCTGCCGTT
GCGTCTTGCGCATCGCAGGGTGGATAAGCAGTGGTTTGAACACATCACCAGTGTGTTGAAGATTGATAATCGTATGC
GCCATTTGCCTGGGGAGCTGTCTGGCGGTCAGCAACAACGCGCCGCGATTGCCCGGGCGTTGATGTCTAGGCCCGAT
ATTGTCATTGCGGATGAGCCAACAGGAAGTTTGGATTCCGTCACCAGCGATGCAGTGTTGAATTTGTTCCGCAGCAT
TGTTGATGATTTTGGGCAGTCACTTGTGTTTGTCACCCACGATAAAGATGCTGCTCACCGTGGTGACGTGTTGATCA
CAATGCGTGATGGCAAGATCATCGATACGGCAGATTTGCGGGTGGGGCGT
>RXA00410-nucleotide sequence C: downstream
TAATGTTCAGGCTTGCTTTCGCT

>>RXA00526-amino acid sequence
(1-690, translated) 230 residues
MSLIEMRNIV KTYNIGSEGE LTVLHGVDFH VDRGEFVSVV GTSGSGKSTM MNIIGLLDKP TDGTYTLDGV
DVLDISDDAL ASHRAKSIGF VFQNFNLIGR IDALKNVEMP MMYAGIPAKQ RRSRAVELLE MVGMGERLNH
EPNELSGGQK QRVAIARALA NDPEIILADE PTGALDSATG RMVMDIFHQL NKEQGKTIVF ITHNPELADE
SDRVVTMVDG RIIGSEVKHS
>RXA00526-nucleotide sequence A: upstream
GGTGGAGCAGGCGGCGGCTCCTTTTAGTCCTGCGGCCCCTTTTGACCCTGCAGCCCCTGCCGTTTCTGCCAAGCAAA
CCGTGGGCCAGGTGATTTAGCCT
>RXA00526-nucleotide sequence B: coding region
ATGAGCCTCATCGAAATGCGAAATATTGTCAAGACCTACAACATTGGATCTGAAGGTGAACTCACCGTGTTGCACGG
TGTGGATTTCCATGTGGACCGTGGCGAATTCGTGTCGGTTGTGGGTACGTCCGGCTCAGGTAAATCAACGATGATGA
ACATCATTGGGTTGTTGGATAAGCCAACTGATGGCACGTACACCTTGGATGGCGTGGATGTGTTGGATATCAGCGAT
GATGCTTTGGCGAGCCACCGCGCTAAATCGATTGGTTTTGTGTTTCAGAACTTCAATCTGATTGGCCGGATCGATGC
GTTGAAGAATGTGGAAATGCCCATGATGTATGCGGGCATTCCGGCTAAGCAGCGGAGAAGTCGTGCGGTTGAATTAT
TGGAAATGGTCGGGATGGGTGAGCGTCTCAACCATGAGCCCAATGAGCTTTCGGGTGGTCAGAAGCAGCGCGTGGCC
ATTGCTCGCGCGTTGGCGAACGATCCTGAGATCATTCTTGCTGATGAACCAACTGGTGCGTTGGATTCTGCAACGGG
CCGGATGGTGATGGATATTTTCCACCAGCTCAACAAGGAGCAGGGCAAAACCATCGTGTTTATTACTCACAACCCTG
AGCTTGCTGATGAATCTGATCGGGTGGTCACCATGGTTGACGGGCGCATCATTGGGTCTGAGGTGAAACACTCA
>RXA00526-nucleotide sequence C: downstream
TGAGCCTTGCAGAATCAATTCTT

>>RXA00735-amino acid sequence
(1-669, translated) 223 residues
GPTGAGKTTL VNLIMRFYDI NSGSITLGET AQDAVDIRTM AREDLRSRTG MVLQDTWLFA GTIRDNILYG
RPEATEEEML AASKAAYVDR FVRSLPEGYD TVLDDEAMNL SVGERQLITI ARAFLANPRL LILDEATSSV
DTRTELLIQR AMSKLRQDRT AFVIAHRLST IRDANLILMM KDGEIVEQGN HRELMALEGA YWELYNSQFN
APAKEELQAD GDH
>RXA00735-nucleotide sequence B: coding region
GGTCCCACCGGTGCGGGCAAGACCACATTGGTGAATCTGATCATGCGTTTCTACGACATCAACAGCGGTTCCATCAC
TCTTGGTGAAACAGCACAAGACGCCGTGGATATCCGCACCATGGCTAGAGAAGATCTGCGATCACGAACCGGCATGG
TGTTGCAGGATACGTGGCTGTTTGCCGGAACCATCAGGGATAACATTCTTTACGGTAGACCTGAAGCAACTGAGGAA
GAAATGCTTGCTGCGTCCAAGGCCGCCTACGTGGATCGTTTTGTCCGTTCCCTGCCAGAAGGCTACGACACCGTACT
TGATGATGAAGCCATGAACCTATCGGTGGGTGAACGCCAGCTGATCACCATCGCGCGTGCATTCTTGGCTAATCCCC
GACTGCTGATTCTGGATGAAGCCACCTCATCGGTGGATACGCGTACCGAATTGTTGATTCAGCGCGCCATGTCCAAG
CTGCGCCAAGACCGCACCGCCTTCGTCATCGCGCACCGGTTGTCCACGATTCGTGATGCCAACCTGATTTTGATGAT
GAAAGACGGCGAGATCGTGGAGCAGGGCAATCACCGTGAGTTGATGGCCCTGGAGGGCGCATATTGGGAGTTGTATA
ACTCCCAATTCAACGCCCCCGCGAAAGAAGAATTACAGGCTGACGGAGATCAC
>RXA00735-nucleotide sequence C: downstream
TGATGATTTCTTCTTAGGCTTTC

>>RXA00734-amino acid sequence
(1-453, translated) 151 residues
RHLRYGNEDA TETQLWQALA IAQAADFVRE MPEGLDSEIA QGGTNVSGGQ RQRLAIARAL LKQPEIYIFD
DSFSALDVST DAALRRALST NLPDATKLIV AQRVSTIRDA DQIVVLDNGE VVGIGTHTNL LNTCGTYREI
VESQETAQAQ S
>RXA00734-nucleotide sequence B: coding region
AGGCACCTGCGTTATGGCAATGAAGATGCCACGGAAACGCAGCTGTGGCAGGCGCTTGCAATTGCTCAGGCGGCGGA
CTTTGTGCGTGAGATGCCAGAGGGTCTTGATTCTGAGATTGCTCAGGGTGGAACCAATGTTTCTGGTGGTCAGCGCC
AGCGACTAGCCATTGCCAGGGCGTTGTTGAAGCAACCTGAGATCTATATTTTCGACGATTCTTTCTCCGCCCTCGAT
GTGAGCACAGACGCCGCTCTTCGCCGAGCGCTGAGCACCAACCTGCCGGATGCAACCAAGTTGATTGTCGCCCAGCG
TGTCAGCACGATTCGAGATGCCGATCAGATTGTGGTGCTTGATAACGGCGAGGTTGTCGGTATTGGAACGCACACGA
ATTTGCTGAACACGTGCGGTACCTACCGTGAAATTGTTGAATCCCAAGAGACTGCGCAGGCGCAATCA
>RXA00734-nucleotide sequence C: downstream
TGAGTAATACTGCAGGCCCCCGC

>>RXA00733-amino acid sequence
(1-408, translated) 136 residues
MSNTAGPRGR SHQADAAPNQ KAQNFGPSAK RLFGILGHDR NTLIFVIFLA VLSVGLTVLG PWLLGKATNV
VFEGFLSKRM PAGASKEDII AQLQAAGKHN QASMMEDMNL VPGSGIDFEK LAMILGLVIG AYLIRS
>RXA00733-nucleotide sequence A: upstream
ACGGCGAGGTTGTCGGTATTGGAACGCACACGAATTTGCTGAACACGTGCGGTACCTACCGTGAAATTGTTGAATCC
CAAGAGACTGCGCAGGCGCAATC
>RXA00733-nucleotide sequence B: coding region
ATGAGTAATACTGCAGGCCCCCGCGGGCGTTCCCATCAGGCAGACGCCGCGCCGAATCAAAAGGCACAGAATTTCGG
ACCATCTGCCAAAAGGCTTTTCGGAATTCTAGGCCATGACCGTAACACCTTAATTTTTGTTATCTTCCTAGCCGTCC
TGAGCGTTGGACTTACCGTCTTGGGCCCATGGTTGCTGGGTAAAGCCACCAACGTGGTGTTTGAAGGATTCCTATCT
AAGCGCATGCCGGCTGGTGCGTCAAAGGAAGATATCATCGCGCAGTTGCAGGCTGCAGGTAAACATAATCAGGCTTC
CATGATGGAAGACATGAACCTTGTTCCAGGCTCAGGCATTGATTTTGAAAAATTAGCCATGATCCTCGGACTGGTGA
TCGGTGCTTATCTCATTCGTAGC

>>RXA00878-amino acid sequence
(1-1863, translated) 621 residues

MRLLGRILKT TSALWPYYLG IIVVSIVIAA LSLLSPFILR EATDSIVSAV TGSNTVDAVT RTIIFLALAL
FVASFLNTVM TNIGGYIGDV MASRMRQILA TRYYAKLLAL PQKYFDNQVT GTIIARLDRS INGITQFMQS
FSNNFFPMLI TMVAVLIISA IFYWPLAILL AMLFPIYMWL TALTSKRWQK YEGEKNHEID VANGRFAEVV
GQVKVVKSFV AETRELADFG GRYGKTVAIT RPQSGWWHRM DTLRGAALNI IFLAIHLLIF YRTLHGHFTI
GDMVMLIQLV TMAQQPVYMM SYIVDSAQRA IAGSRDYFEV MAQQVEPTAN KELVDATLAS DTPRISVGTP
AALPAGEPAM EFKNVTFAYE EGKPVISDVS ITARHGERIA LVGESGGGKS TLVNLLLGLY KPNSGSLAVC
GVDVKDLTSE ELRASVGVVF QDASLFSGSI AENIAYGRPG ATREEIIEVA KKANAHEFIS AFPEGYETVV
GERGLKLSGG QKQRVSVARA MLKDAPLLVL DEATSALDTK SEQAVQAGLE QLMENRTTLM IAHRLSTIAG
VDTIVTIQNG RVEEVGSPTE LAVSGGIYSE LLRLTNSTAE ADRERLRAFG FTGDAPAEEE D

>RXA00878-nucleotide sequence A: upstream

CGAGATTAGGTCCGCTTCAGTTGTGGTGGCTCCGAATCTGATGAACAATGATCATTCCTAATTCATTTACATCTTTA
TCAAAGAGAGCCACCACCTACTA

>RXA00878-nucleotide sequence B: coding region

ATGCGACTTCTTGGTCGAATTTTAAAAACCACGTCTGCGCTTTGGCCCTACTATCTCGGAATTATCGTCGTATCCAT
TGTGATCGCGGCGTTGTCGCTGCTGTCGCCGTTTATTCTCCGCGAAGCAACAGATTCCATTGTTTCTGCAGTAACCG
GATCTAACACCGTCGATGCAGTTACTCGCACTATTATTTTCTTAGCTTTAGCCCTGTTTGTCGCAAGCTTCCTCAAT
ACGGTGATGACCAACATCGGTGGCTACATCGGTGATGTCATGGCATCTCGTATGCGCCAGATTCTGGCCACGCGCTA
TTACGCAAAGCTGTTGGCGCTGCCTCAGAAGTATTTTGATAATCAGGTCACCGGCACCATCATCGCCCGCCTTGATC
GATCAATCAACGGCATCACGCAGTTCATGCAGAGCTTCTCCAACAACTTCTTCCCCATGCTCATCACCATGGTGGCA
GTGCTGATTATTTCCGCGATTTTCTACTGGCCTCTGGCAATTCTGCTGGCCATGTTGTTCCCGATTTACATGTGGCT
GACGGCGTTGACATCGAAACGCTGGCAGAAATATGAGGGCGAGAAAAACCATGAAATCGACGTGGCTAACGGCCGCT
TCGCTGAGGTTGTCGGCCAGGTCAAGGTTGTTAAATCATTCGTCGCAGAGACCCGCGAGCTGGCTGATTTCGGTGGG
CGTTACGGCAAAACAGTAGCGATTACCCGGCCGCAATCCGGTTGGTGGCACCGCATGGATACTCTCCGTGGCGCGGC
ACTAAATATCATCTTCCTGGCCATTCACCTGCTGATTTTCTACCGCACCTTGCACGGCCATTTCACCATCGGCGACA
TGGTCATGCTCATCCAGCTTGTCACCATGGCGCAGCAACCGGTGTACATGATGAGCTACATCGTCGACTCCGCGCAG
CGCGCCATCGCCGGCTCCCGCGACTACTTCGAGGTCATGGCGCAGCAGGTCGAGCCCACCGCCAATAAGGAGCTTGT
CGACGCCACCCTCGCCTCAGACACTCCACGCATCAGTGTGGGCACGCCCGCCGCGCTGCCCGCTGGAGAACCAGCGA
TGGAATTCAAAAACGTCACCTTCGCCTACGAAGAAGGCAAGCCGGTTATTTCCGACGTGTCCATTACCGCCCGCCAC
GGCGAGCGCATCGCGTTGGTCGGTGAATCCGGCGGCGGTAAATCCACCCTGGTCAACCTTCTGTTAGGTCTGTACAA
ACCAAACAGCGGCAGCCTTGCAGTATGTGGCGTGGATGTTAAAGATCTGACTTCCGAGGAACTTCGCGCATCCGTGG
GTGTGGTCTTCCAGGACGCCAGCTTGTTCTCTGGATCTATTGCAGAAAACATCGCCTACGGTCGCCCAGGTGCCACC
CGCGAAGAGATCATCGAAGTGGCTAAGAAAGCCAACGCACATGAGTTCATTTCCGCCTTCCCTGAAGGATATGAAAC
CGTCGTCGGTGAACGCGGACTCAAACTTTCTGGTGGCCAGAAGCAGCGCGTCTCTGTGGCACGGGCCATGCTTAAAG
ATGCCCCACTTCTTGTTCTCGATGAAGCCACCTCTGCACTGGATACCAAGTCTGAGCAGGCAGTCCAAGCCGGTTTG
GAACAGCTGATGGAAAACCGCACCACCTTAATGATCGCCCACCGCCTGTCCACCATCGCAGGCGTCGATACCATCGT
GACCATCCAAAACGGACGGGTTGAAGAGGTCGGATCTCCTACCGAGCTCGCAGTCTCAGGCGGTATCTATTCCGAAC
TGCTGCGCCTGACCAACTCCACAGCAGAAGCCGACCGGGAGCGTCTGCGCGCCTTTGGTTTCACTGGCGATGCACCA
GCTGAAGAAGAGGAC

>RXA00878-nucleotide sequence C: downstream

TAGCCCCGCGAAAGAACAATCCC

>>RXA01191-amino acid sequence
(1-1407, translated) 469 residues
VSLDANTIET AGRGDVISRI ADDSREVSTA ASTVVPLMVQ AGFTVVISAF GMAAVDWRLG LVGLVAIPLY
WTTLRVYLPR SGPLYTRERE AFGVRTQRLV GAVEGAETLR AFRAEDTELK RIDAASGEAR DISISVFRFL
TWAFSRNNRA ECITLVLILG TGFYLVNIDL VTVGAVSTAA LIFHRLFGPI GTLVGMFSDI QSASASLIRM
VGVINAASNQ VSGTSPASAS TALTLFDVSH HYHTAPVIKN ASVQLEPGEH IAIVGATGAG KSTLALIAAG
LLSPTSGQVA LGGSSFSNVE PEALRQKIAM VSQEIHCFRG SVLDNLRIAR PEATDADIHA VLADIGDSWL
ERLPQGIDTI VGDGAFRLTS VENQIMALAR VHLADLAIVI LDEATAESGS DHAKQLEDAA LKVTENRSAI
IVAHRLNQAK TADRIIVMDS GEIIESGTHE ELRAIGGRYE QLWTAWSAR
>RXA01191-nucleotide sequence B: coding region
GTGAGTTTGGATGCGAACACGATTGAAACGGCGGGGCGCGGCGACGTGATTTCGCGTATCGCGGATGATTCGCGGGA
GGTGTCCACTGCGGCGAGCACCGTGGTGCCGCTGATGGTGCAGGCGGGCTTTACCGTGGTGATTTCCGCGTTTGGCA
TGGCGGCGGTTGATTGGCGCCTCGGCCTTGTCGGTTTGGTCGCGATCCCGCTGTATTGGACCACGTTGCGCGTCTAT
TTACCCCGCTCAGGTCCGCTTTATACGCGTGAGCGCGAGGCCTTTGGGGTGCGCACGCAGCGGCTTGTCGGCGCAGT
CGAAGGCGCGGAAACCTTGCGCGCTTTCCGCGCAGAAGATACAGAATTAAAGCGTATCGACGCAGCCTCCGGCGAAG
CCCGCGACATTTCCATTTCTGTTTTCAGGTTCCTCACATGGGCATTTTCCCGCAACAACCGCGCGGAATGCATCACC
CTCGTGCTCATCTTGGGCACCGGCTTTTACCTGGTCAACATCGATCTGGTCACCGTCGGCGCAGTCTCAACCGCCGC
ACTGATCTTCCACCGACTCTTCGGTCCAATCGGCACGCTCGTGGGCATGTTCTCCGACATCCAATCCGCCAGCGCAT
CGCTGATCCGCATGGTGGGCGTTATTAACGCGGCATCGAACCAGGTCAGCGGCACCTCGCCGGCGTCTGCCAGCACC
GCTTTAACGCTTTTCGACGTCTCCCACCACTATCACACTGCACCCGTCATCAAGAATGCATCCGTGCAGCTGGAACC
AGGGGAACACATCGCCATTGTGGGTGCGACCGGCGCTGGTAAAAGCACGCTCGCCCTCATTGCGGCAGGCCTGCTCA
GCCCAACTTCCGGGCAGGTGGCTCTCGGCGGATCGAGTTTTTCTAACGTCGAACCGGAAGCATTGCGCCAGAAGATC
GCGATGGTCAGCCAAGAAATCCACTGCTTCCGAGGATCTGTTTTAGATAATCTTCGTATCGCACGCCCCGAAGCCAC
CGATGCGGACATCCACGCCGTTCTCGCCGATATTGGTGATTCCTGGTTGGAGCGCTTACCGCAAGGCATAGACACCA
TCGTGGGTGATGGCGCTTTCCGTTTAACCTCTGTGGAAAACCAGATCATGGCGCTTGCTCGCGTACATTTGGCCGAC
CTAGCAATCGTCATCCTTGATGAAGCAACGGCTGAATCAGGCTCTGATCATGCAAAACAGCTTGAAGATGCAGCCCT
TAAAGTCACTGAAAACAGATCAGCCATCATCGTGGCTCACCGCCTCAACCAAGCGAAAACCGCCGATCGCATCATCG
TCATGGACTCCGGAGAAATCATAGAATCTGGAACCCATGAAGAGCTTCGAGCGATCGGCGGTCGATATGAACAACTG
TGGACTGCGTGGTCTGCGCGC
>RXA01191-nucleotide sequence C: downstream
TAATTAGCCACCCAAGACCACGC

>>RXA01212-amino acid sequence
(1-813, translated) 271 residues
GLNFHVQRGE VFGLLGTNGA GKTSTLEVIE GLSAPSSGTV RISGLDPVAD RAILRPELGI MLQSGGLPSQ
LTVAETMDMW HGTCTYPRAI KDVLADVDLL HRENVKVGAL SGGEQRRLDL ACALLGDPSI LFLDEPTTGL
DPESRRHTWQ LLLDLKQRGV TMMLTTHYLE EAEFLCDRIA IMNAGEIAVE GTLDELVARE KSIISFVLRG
GQVELPVLSG AEIIRDNNHV RIATTTLQQH TLEILTWAAE TGIALEGFAA KPATLESVFM D
>RXA01212-nucleotide sequence B: coding region
GGCCTGAATTTCCATGTACAGCGCGGTGAAGTATTTGGTCTGCTCGGCACCAACGGGGCCGGCAAAACCTCCACCTT
GGAAGTCATCGAAGGACTTTCCGCACCCAGCTCCGGCACCGTGCGCATCTCCGGGCTTGACCCCGTTGCCGACCGCG
CGATCCTGCGCCCCGAGCTCGGCATCATGCTGCAATCAGGCGGCCTGCCATCACAGCTCACCGTCGCCGAAACCATG
GACATGTGGCACGGCACCTGCACGTATCCGCGCGCCATTAAAGATGTGCTTGCCGACGTCGACCTCCTACACCGCGA
AAACGTCAAGGTCGGCGCGCTTTCCGGAGGCGAACAACGACGCCTTGATTTGGCCTGCGCACTGCTTGGCGACCCCT
CAATTTTGTTCCTCGACGAACCCACCACCGGCCTCGACCCAGAATCTAGGCGCCACACCTGGCAACTCCTGCTGGAC
CTGAAACAGCGCGGCGTCACCATGATGCTGACCACCCACTACCTGGAGGAAGCCGAATTCCTCTGCGACCGGATTGC
CATCATGAACGCCGGTGAGATCGCAGTGGAAGGCACCTTGGATGAACTGGTGGCCCGCGAGAAGTCGATCATCAGTT
TCGTGCTGCGTGGCGGGCAGGTGGAGTTGCCGGTCTTGAGTGGGGCTGAAATCATCCGCGACAACAACCACGTCCGC
ATCGCCACCACCACCCTGCAGCAGCACACCTTAGAAATACTTACCTGGGCTGCAGAGACCGGGATCGCGCTGGAAGG
CTTCGCTGCAAAACCCGCCACCTTGGAATCCGTATTCATGGAC

>>RXA02095-amino acid sequence
(1-1404, translated) 468 residues
MKTEQSQKAQ LAPKKAPEKP QRIRQLISVA WQRPWLTSFT VISALAATLF ELTLPLLTGG AIDIALGNTG
DTLTTDLLDR FTPSGLSVLT SVIALIVLLA LLRYASQFGR RYTAGKLSMG VQHDVRLKTM RSLQNLDGPG
QDSIRTGQVV SRSISDINMV QSLVAMLPML IGNVVKLVLT LVIMLAISPP LTIIAAVLVP LLLWAVAYSR
KALFASTWSA QQKAADLTTH VEETVTGIRV VKAFAQEDRE TDKLDLTARE LFAQRMRTAR LTAKFIPMVE
QLPQLALVVN IVGGGYLAMT GHITVGTFVA FSSYLTSLSA VARSLSGMLM RVQLALSSVE RIFEVIDLQP
ERTDPAHPLS LPDTPLGLSF NNVDFRGILN GFELGVQAGE TVVLVGPPGS GKTMAVQLAG NFYQPDSGHI
AFDSNGHRTR FDDLTHSDIR RNLIAVFDEP FLYSSSIPRE HLDGFGCQ
>RXA02095-nucleotide sequence A: upstream
CTCTCTTGGTCCTCTCCCCACCCATTTTTAAGTACTCAAGACCCTTCCAACAGAAAGGATTACTCCCCCAACAGGCT
CAAAAATACTGAAAGGCTCACGC
>RXA02095-nucleotide sequence B: coding region
ATGAAAACTGAGCAATCCCAAAAAGCACAATTAGCCCCTAAGAAAGCACCTGAAAAGCCACAACGCATCCGCCAACT
TATTTCCGTGGCGTGGCAGCGACCTTGGCTCACCTCATTCACCGTAATCAGCGCTTTAGCTGCAACGTTGTTTGAAC
TTACACTTCCTCTTTTGACCGGTGGCGCCATCGATATCGCGCTCGGAAATACCGGAGATACTTTAACCACTGACCTG
CTGGACCGGTTCACTCCGAGTGGATTAAGCGTGTTGACCAGCGTCATTGCCCTTATCGTGCTTCTCGCGTTGCTTCG
CTATGCCAGTCAATTTGGACGGCGATACACCGCAGGCAAGCTCAGCATGGGGGTACAGCATGATGTCCGGCTTAAAA
CGATGCGCTCATTGCAGAACCTCGATGGGCCAGGTCAGGACTCTATTCGCACAGGCCAAGTAGTCAGTCGGTCCATT
TCGGATATCAACATGGTGCAAAGCCTTGTGGCGATGTTGCCGATGTTGATCGGAAATGTGGTCAAGCTTGTGCTCAC
TTTGGTGATCATGCTGGCTATTTCCCCGCCGCTGACCATCATCGCTGCAGTGTTGGTGCCTTTGCTGTTGTGGGCCG
TGGCCTATTCGCGAAAAGCGCTTTTTGCGTCCACGTGGTCGGCCCAGCAAAAGGCTGCGGATCTGACCACTCATGTG
GAAGAAACTGTCACGGGTATCCGCGTGGTCAAGGCATTTGCGCAGGAAGACCGCGAGACCGACAAATTGGATCTCAC
CGCACGTGAGTTATTTGCCCAGCGCATGCGCACTGCACGTCTGACGGCAAAGTTCATCCCCATGGTTGAGCAGCTTC
CGCAGCTTGCTTTGGTGGTCAACATTGTTGGCGGTGGCTATTTGGCCATGACTGGTCACATCACGGTGGGCACGTTT
GTGGCGTTTTCTTCCTATCTCACTAGCTTGTCGGCGGTGGCTAGGTCCCTGTCGGGCATGCTCATGCGCGTGCAGTT
GGCGCTGTCTTCTGTGGAGCGCATCTTTGAAGTCATTGATCTTCAGCCTGAACGCACCGATCCTGCACACCCCCTGT
CACTTCCCGACACTCCCCTGGGTCTGTCGTTCAACAACGTAGATTTCCGTGGGATTCTCAACGGTTTTGAGCTGGGT
GTTCAGGCCGGTGAAACCGTTGTGTTGGTGGGCCCTCCAGGTTCAGGCAAGACCATGGCTGTGCAGCTTGCTGGAAA
CTTTTATCAACCAGACAGCGGCCACATCGCCTTTGATAGCAACGGCCATCGCACTCGCTTCGACGACCTCACCCACA
GCGATATCCGCAGGAATCTCATCGCGGTTTTTGATGAGCCGTTCTTGTACTCCTCCTCCATACCGCGAGAACATCTC
GATGGGTTTGGATGTCAG
>RXA02095-nucleotide sequence C: downstream
TGATGAGCAGATCGAACACGCAG

>>RXA02074-amino acid sequence
(1-1623, translated) 541 residues
MRSLLRDIPA VGWLITATIV VRTLVVALVI VGIGLLIDVP SPAHSAMLWW VLAGATAAAA LLCAEAVLPQ
RIRARVERSW RRQLAAKNLE LNSSSSDDAQ LITLATEATS KASTYTVMFL GPYFAVFLAP LTVIAVVGAA
ISWPIAGILC LGLCVIPFVI SWAQRMLKGA GAGYGRASGQ LAGVFLESVR TLGTTMMLNA AGQRRQIITQ
RAENMRSQVM SLLYRNQLMI LVTDGVFGVA TTMVAAVFAI GGFFSGSLTL GQAVALVLLA RLLIDPINRM
GRTFYTGMAG KPSLIAIEKA LATTFTDQPT QQGQRHDGDL VVNNLKIARD HRDIVHGISF SIPRGSHIAV
VGPSGAGKSS VALALSGLLE FDGAISLGGH NCEMLDLRAS VSFVPQSPTL FSGSIKSNID LARTGVDSDH
IHAALLGEEL PADLKVGETG KGVSGGQAAR ISIARGLVKN AAVIVLDEAT AQLDYTNARQ VRHLAKSLEC
TLVEITHRPS EALDADFIIV LEDGQLTMMD TPSNVSQHNA FFRTAVMEEE Q
>RXA02074-nucleotide sequence A: upstream
CGGGGGAAGGCCGTGTCGCATGCTCGGGCTAGCCTTGGATCTCAAGAAGAATTCGACTGGTTTAAAGTCTGGGCTTT
AAGTGCAGAAAGGTTGTGGATTG
>RXA02074-nucleotide sequence B: coding region
ATGCGCTCCCTGCTTCGTGATATCCCTGCGGTGGGTTGGCTAATCACCGCGACGATTGTTGTGCGCACGCTCGTTGT
TGCGCTGGTCATCGTTGGGATCGGCTTGCTTATCGACGTCCCCTCGCCCGCTCATTCAGCCATGTTGTGGTGGGTTC
TGGCAGGTGCCACGGCAGCAGCTGCGCTGCTGTGCGCGGAAGCGGTGCTCCCCCAACGTATTCGTGCACGAGTTGAA
CGATCCTGGCGGCGGCAGTTGGCTGCTAAAAATCTGGAGCTGAATTCCAGTTCGTCAGATGATGCCCAGTTGATCAC
ACTGGCAACTGAAGCCACCTCAAAAGCATCCACTTACACAGTGATGTTTCTGGGGCCTTACTTTGCAGTATTTTTGG
CCCCACTGACAGTTATTGCCGTTGTCGGCGCGGCTATTTCCTGGCCGATTGCGGGGATACTGTGCCTCGGGTTGTGC
GTGATACCTTTCGTTATTTCTTGGGCACAGCGCATGTTGAAAGGCGCTGGCGCGGGATACGGGCGAGCATCTGGGCA
GTTGGCAGGCGTGTTTTTGGAATCGGTGCGCACACTAGGCACCACGATGATGCTGAATGCCGCTGGGCAGCGCAGGC
AGATCATCACACAGCGCGCAGAGAATATGCGCTCCCAAGTGATGTCATTGCTGTACCGAAATCAGTTGATGATTCTG
GTGACCGACGGCGTGTTTGGAGTTGCCACCACAATGGTTGCTGCGGTGTTTGCCATTGGAGGATTCTTTTCAGGCTC
TCTTACTCTCGGCCAAGCTGTAGCACTCGTATTGCTGGCCAGGCTGCTTATTGATCCCATCAACCGCATGGGTCGCA
CGTTTTACACCGGCATGGCAGGCAAACCCTCGCTGATCGCCATTGAAAAAGCCCTCGCGACAACCTTTACTGATCAG
CCAACTCAACAGGGACAGCGCCACGATGGGGATCTGGTGGTCAACAACTTGAAGATCGCCCGCGATCACAGGGACAT
TGTGCACGGTATCTCTTTCAGCATTCCCCGCGGTTCCCACATCGCGGTGGTAGGTCCCAGTGGCGCTGGTAAATCCT
CTGTGGCTCTAGCGTTGTCCGGACTTTTAGAGTTTGATGGTGCGATTTCCCTCGGCGGCCACAACTGTGAGATGTTA
GATCTTCGCGCCTCAGTCAGTTTCGTGCCCCAATCCCCCACGCTGTTTAGCGGAAGCATCAAAAGCAATATCGATCT
GGCGCGCACGGGTGTTGATTCTGATCACATCCACGCAGCACTTTTAGGCGAAGAACTCCCCGCGGACCTCAAAGTCG
GTGAAACCGGCAAAGGTGTCTCCGGCGGCCAAGCAGCACGCATTTCCATTGCCCGAGGTTTAGTAAAGAATGCTGCC
GTGATTGTTCTCGACGAGGCGACCGCACAACTCGACTACACCAACGCCCGCCAGGTTCGACATCTTGCCAAATCCCT
TGAGTGCACGTTGGTTGAGATCACCCACCGCCCATCAGAAGCCCTCGATGCAGACTTCATCATTGTTTTAGAGGATG
GCCAATTGACCATGATGGATACACCCAGCAACGTTTCCCAGCACAATGCGTTTTTCCGCACCGCTGTGATGGAGGAA
GAACAA
>RXA02074-nucleotide sequence C: downstream
TGATTTCCCGACTTCTCCAATTG

>>RXA02224-amino acid sequence
(1-1797, translated) 599 residues
MAQHERVADA LQPASLAESW RELKTMPSGP KAWWYVSFVV ISVVTVVAMV GTSNLLGYSV DLINGQSLPL
IGSGSTAMIW LLGLVGAGIL AETAGRALLQ LVINTLARRL SVDLRKAALS SALRAPVPDV MELGTGNVIS
RLTQDIDNTV RIVGMVGVRL VITILILPSS LFALMTIHWT FVILFIAVIV VLIPSGRKAV RAIPSATNIV
SSTEARRNNL LLDTIRGIET LRVLKLGAWG VQRMRQASWT AVQATADRAP IFTRLLALGS IAYGLLLIGV
FGLSAFWVAQ DAMSIGAATA AVFVVVRMEI HVFNVLFFAS EIQSASTSLG RAVSLAQMAR RTEQLSESAD
CTEPPSVTVQ DVTFKYPGGV AILEDFNLVL EAGTTTALVG TSGAGKSTLA GVIAGLQRPD SGAVLVGGIN
TATVTDTWTT RQVALISQEV HLFAGTLAED LRMANAHATD AQLHAALESV GLGQMTTAFR RFFPSGLDTK
IGAGAEELTP EIQQQISLAR IVLRNPPVLI MDEATSEAGS DDARMLEKAA TEIARNRTTL VVAHRLDQAV
VADRIIVMEQ GTITEDGTHQ ELLAFEGRYA QLYQRWSAQ
>RXA02224-nucleotide sequence A: upstream
GCTTCGTCGAGGCCGGAAAACCGTCGTCATTACGTCGAACCCGACGTGGCACGGCGTCGCAAAGCAGATGCAATCTG
ATTTTTCGGAAGGGGTGAAGTAG
>RXA02224-nucleotide sequence B: coding region
ATGGCGCAGCATGAGCGCGTTGCGGATGCGCTGCAGCCGGCGTCGTTGGCGGAGTCGTGGCGTGAGCTGAAAACGAT
GCCTTCGGGGCCCAAGGCCTGGTGGTATGTGAGTTTCGTGGTTATTAGCGTGGTCACGGTCGTGGCGATGGTCGGCA
CGTCCAACTTGTTGGGCTATTCCGTTGATCTGATCAATGGGCAGTCGTTGCCGCTGATCGGTTCAGGATCGACCGCA
ATGATCTGGTTGCTTGGTTTGGTGGGCGCTGGAATTTTAGCAGAAACTGCCGGTCGCGCGCTGCTGCAATTGGTGAT
CAACACCTTGGCACGTCGCCTGTCGGTGGATCTGCGGAAAGCTGCGCTGTCTTCGGCGTTGCGTGCACCGGTTCCTG
ATGTCATGGAATTGGGCACGGGAAACGTGATTAGCCGCCTGACGCAAGACATCGATAACACTGTGCGCATCGTCGGC
ATGGTAGGTGTGCGTTTGGTGATCACCATTTTGATTCTGCCCAGCTCCTTGTTCGCGTTGATGACCATTCACTGGAC
CTTTGTGATCCTGTTCATCGCAGTGATTGTGGTGCTGATTCCCAGCGGTCGGAAAGCCGTGCGAGCTATTCCTTCGG
CAACAAATATTGTGTCCAGTACGGAGGCGCGTCGAAACAATCTGCTCCTCGATACGATCCGTGGCATTGAAACACTG
CGTGTGCTCAAGCTCGGTGCGTGGGGTGTGCAGCGGATGCGCCAAGCGTCGTGGACTGCGGTGCAAGCAACAGCTGA
TCGCGCGCCGATTTTCACTCGTCTGCTCGCCCTTGGTTCGATTGCTTATGGCCTGCTGCTAATTGGCGTGTTTGGGC
TCAGTGCGTTTTGGGTTGCCCAGGATGCGATGAGCATTGGAGCGGCAACGGCAGCAGTTTTCGTGGTTGTGCGCATG
GAAATTCACGTGTTCAACGTGCTGTTCTTCGCATCGGAAATTCAGAGTGCGTCTACTTCTCTTGGTCGCGCGGTGTC
CCTTGCCCAGATGGCTCGTCGCACCGAACAGCTGTCTGAGTCTGCCGATTGCACAGAACCACCCTCCGTGACTGTGC
AGGACGTGACGTTTAAATATCCCGGCGGCGTGGCCATTTTGGAGGATTTCAATCTGGTCTTGGAAGCAGGAACAACC
ACAGCGCTGGTCGGTACTTCTGGTGCGGGAAAATCCACGCTCGCGGGCGTCATTGCGGGGCTGCAGCGCCCTGATTC
CGGCGCCGTTTTGGTCGGGGGCATCAACACCGCCACCGTCACCGACACGTGGACTACCCGCCAGGTTGCGCTGATCA
GCCAGGAAGTCCACCTTTTCGCAGGCACTCTGGCAGAGGATCTTCGCATGGCCAATGCGCACGCCACCGACGCGCAG
CTCCACGCAGCGCTCGAGTCGGTCGGGCTCGGGCAAATGACAACTGCTTTTCGACGTTTCTTTCCATCCGGATTAGA
CACCAAAATTGGCGCCGGCGCAGAAGAACTCACCCCTGAAATCCAACAGCAAATCTCTCTTGCCCGCATCGTGCTCC
GCAATCCACCTGTGTTGATCATGGATGAAGCCACCAGTGAAGCCGGCAGCGATGATGCCCGCATGTTGGAAAAAGCC
GCCACAGAAATCGCACGAAACCGCACCACCTTGGTTGTTGCGCACCGCCTTGACCAAGCAGTTGTCGCAGATCGCAT
CATCGTGATGGAACAAGGCACAATCACCGAAGACGGCACTCACCAGGAATTACTTGCTTTTGAGGGCCGCTACGCGC
AGCTGTATCAACGATGGAGTGCTCAA
>RXA02224-nucleotide sequence C: downstream
TAGTTCAAATCCACCACAAACTC

>>RXA01947-amino acid sequence
(1-279, translated) 93 residues
ILLDEPTNDL DVETLGSLEN ALQNFPGCAV VISHDRWFLD RTCTHILAWE GNIAEGQWYW FEGNFEDYEK
NKVERLGADA ARPSRVTHRK LTR
>RXA01947-nucleotide sequence B: coding region
ATCCTCCTCGATGAGCCTACAAACGACCTCGACGTGGAAACTCTGGGCTCCCTGGAAAATGCACTCCAGAACTTCCC
TGGCTGTGCAGTGGTCATTTCCCACGACCGTTGGTTCCTGGACCGCACCTGTACCCACATCCTCGCGTGGGAAGGCA
ACATCGCGGAAGGCCAGTGGTACTGGTTCGAAGGCAACTTCGAAGACTACGAGAAGAACAAGGTGGAGCGCCTCGGT
GCTGACGCCGCACGTCCTTCCCGTGTCACTCACCGCAAGCTGACTCGC
>RXA01947-nucleotide sequence C: downstream
TAGAGTCGAAGCAGCGCTAAACC

>>RXA01946-amino acid sequence
(1-1275, translated) 425 residues
IRKYSRLEEQ FQSLGGYEAD AEAAQICDNL GLEARILDQQ LKTLSGGQRR RVELAQILFA ATNGSGKSKT
TLLLDEPTNH LDADSITWLR DFLAKHEGGL IMISHDVELL GAVCNKIWYL DAVRSEADVY NMGFSKYVDA
RALDEARRRR ERANAEKKAG ALKDQAARLG AKATKAAAAK QMIARAERMI DNLDEIRVAD RAANIVFPEP
APCGKTPLNA KGLTKMYGSL EVFAGVDLAI DKGSRVVVLG FNGAGKTTLL KLLAGVERTD GEGGIVTGYG
LKIGYFAQEH DTIDPDKSVW QNTIEACADA DQQSLRSLLG SFMFSGEQLD QPAGTLSGGE KTRLALATLV
SSRANVLLLD EPTNNLDPIS REQVLDALRT YTGAVVLVTH DPGAVKALEP ERVIVLPDGT EDLWNDQYME
IVELA
>RXA01946-nucleotide sequence B: coding region
ATCCGCAAGTACTCCAGGCTCGAGGAACAATTCCAGTCGCTCGGCGGCTACGAAGCTGACGCCGAAGCAGCCCAGAT
CTGCGACAACCTCGGCCTCGAGGCACGCATCCTCGACCAGCAGCTTAAAACCCTGTCCGGCGGCCAGCGCCGCCGCG
TCGAGTTGGCGCAGATCCTCTTCGCCGCCACCAACGGCTCCGGCAAATCAAAAACCACATTGCTTCTCGACGAGCCC
ACCAACCACTTGGACGCAGACTCGATCACCTGGCTCCGTGACTTCCTGGCGAAGCACGAAGGTGGACTGATCATGAT
TTCGCACGACGTCGAACTGCTTGGCGCCGTATGTAACAAGATTTGGTACCTCGACGCAGTACGCAGCGAAGCCGATG
TCTACAACATGGGCTTTAGCAAATACGTCGATGCACGTGCACTCGATGAAGCACGCCGACGCCGTGAGCGCGCAAAC
GCCGAAAAGAAGGCCGGAGCCCTCAAGGACCAGGCTGCACGCCTCGGCGCGAAAGCAACCAAGGCTGCCGCAGCTAA
GCAGATGATCGCCCGTGCGGAACGAATGATCGACAACCTCGACGAAATCCGCGTAGCTGACCGCGCCGCCAACATCG
TTTTCCCAGAACCAGCACCCTGTGGAAAAACCCCACTCAACGCCAAGGGCCTGACCAAGATGTACGGCTCCCTCGAA
GTCTTCGCCGGCGTCGACCTAGCCATCGACAAAGGCTCCCGCGTAGTCGTCCTCGGATTCAACGGTGCAGGTAAAAC
CACCCTGCTCAAACTCCTCGCCGGTGTGGAACGCACCGACGGCGAAGGCGGCATCGTCACCGGATACGGCCTCAAAA
TCGGCTACTTCGCCCAGGAACACGACACCATCGACCCCGACAAATCCGTCTGGCAAAACACCATCGAAGCCTGCGCC
GACGCCGACCAACAAAGCCTCCGCAGCCTCCTCGGATCCTTCATGTTCTCCGGCGAACAACTCGACCAACCAGCAGG
AACACTCTCCGGCGGTGAAAAAACCCGCCTCGCACTGGCCACCCTCGTGTCCTCCCGCGCAAACGTCCTGCTTCTCG
ACGAGCCCACCAACAACCTTGACCCGATCTCCCGCGAACAGGTCCTCGACGCACTGCGCACCTACACCGGCGCAGTC
GTCCTGGTTACCCACGACCCGGGTGCAGTCAAGGCCCTTGAGCCAGAACGCGTCATCGTGCTTCCTGATGGCACCGA
GGATCTTTGGAATGATCAGTACATGGAAATCGTGGAATTGGCG
>RXA01946-nucleotide sequence C: downstream
TAGGTTCTAAGGCTGTTTATGCT

>>RXA01881-amino acid sequence
(1-441, translated) 147 residues
MANLINLENV SKTWGLKTLL DGVSLGVQTG DRIGVVGLNG GGKTTLLEVL TGIEKPDQGR VSHNSDLRMA
VVTQRAELND DDTVADVVLG PLGLEVFEWA SNATVRDVLG GLGIVDLGLD TKVGQTFSGG RSADAPTWPP
RWFATLT
>RXA01881-nucleotide sequence A: upstream
ACCGGCCCTGCGGCCTCAACCGCCGACCAGCGCGGCGCACACATTTTGACTGTTTCATAATAAAGACAAACTTAAGT
ATCGGAGTCGAAGAAAAACCACA
>RXA01881-nucleotide sequence B: coding region
ATGGCCAATCTGATTAATCTCGAGAACGTCTCCAAAACCTGGGGATTAAAAACGCTTCTCGACGGTGTCTCCTTAGG
TGTTCAAACCGGCGACCGCATTGGCGTCGTCGGCCTCAATGGTGGCGGAAAAACCACCCTGCTGGAAGTACTTACTG
GCATCGAAAAGCCGGATCAGGGCCGTGTGTCTCACAACTCTGACCTGCGCATGGCTGTGGTGACGCAGCGTGCTGAA
CTCAATGATGACGACACCGTCGCTGACGTGGTGCTTGGACCTTTGGGTTTGGAAGTTTTCGAATGGGCATCAAACGC
CACGGTGCGCGACGTCCTCGGTGGCTTGGGCATTGTCGATCTTGGCCTTGACACCAAGGTGGGGCAAACCTTTTCCG
GTGGGCGAAGCGCCGACGCACCAACCTGGCCGCCGCGCTGGTTCGCGACCTTGACC
>RXA01881-nucleotide sequence C: downstream
TGATCGTGCTCGACGAGCCCACC

>>RXA02253-amino acid sequence
(1-927, translated) 309 residues
MIQSTGVTHT DKSAQENPVK YRDNFTPVII TGMSGAGLST AARVLEDLGW YVAHNIPPQI ILELIDMCAR
EDSPVDKVAV VCDVRSREFR GSLTQVVSEL RDKQLDPTVL FLEARDEVLI KRFDNVRRTH PLQGSQTLQV
GIERERTVLS PVKEDASVVI DTSDLSVHDL RRAIESSFRT IATRTQHVTI ESFGFKHGSP RDADFVVDVR
FLPNPFWVPE LRPFRGVDKP VSDYVLSQKG AEEFLNNFVD MLKDMLPGYR HEGKNFITIG VGCTGGHHRS
VAVSEELAKR IADQTTLDVS VVHRDINRH
>RXA02253-nucleotide sequence A: upstream
TGAGGACATTTCCCAAGTGAAAGGCTTCGGCCCGAAACTTGCGGAGGCTGTCTATGAAGGTCTTCACGCGTCAAAAT
AAGTAGATCGCTAGGATGTAACC
>RXA02253-nucleotide sequence B: coding region
ATGATTCAATCCACTGGGGTCACGCACACTGATAAGTCTGCACAAGAAAATCCTGTGAAGTACAGGGACAATTTCAC
ACCTGTCATCATCACCGGTATGTCAGGCGCAGGTCTGAGCACAGCAGCCCGAGTCCTCGAAGACTTGGGTTGGTATG
TGGCGCACAATATTCCACCACAGATCATCCTGGAATTAATTGATATGTGCGCTCGGGAAGATTCTCCCGTCGACAAA
GTTGCAGTCGTGTGCGATGTGCGCTCCCGTGAATTCCGCGGAAGCCTCACCCAGGTTGTTTCAGAGCTGCGTGATAA
GCAGCTCGATCCCACGGTGTTATTTCTGGAAGCACGCGATGAGGTGCTGATCAAGCGATTCGATAATGTGCGCCGCA
CCCATCCTTTGCAGGGCAGCCAAACCCTTCAGGTGGGTATTGAACGTGAACGAACCGTGCTGTCTCCTGTGAAGGAA
GACGCTTCAGTGGTCATCGATACCTCGGATCTGTCCGTGCATGATTTGCGCCGCGCCATCGAATCCTCGTTTAGGAC
AATCGCCACGCGCACCCAGCACGTCACCATTGAATCATTCGGTTTCAAACACGGCTCACCACGCGACGCCGACTTTG
TTGTGGACGTGCGTTTCTTGCCGAACCCATTCTGGGTTCCAGAGCTGCGCCCATTTAGGGGAGTGGACAAGCCAGTA
TCTGACTATGTGCTCTCCCAAAAAGGCGCAGAAGAATTTTTGAACAACTTTGTGGACATGCTCAAAGACATGCTTCC
GGGATACCGCCACGAAGGAAAAAACTTCATCACAATCGGTGTCGGCTGCACCGGTGGACACCACAGATCAGTTGCGG
TGTCTGAAGAACTAGCCAAAAGAATCGCAGATCAGACCACGCTCGACGTGTCTGTAGTACACCGCGATATTAACCGC
CAC
>RXA02253-nucleotide sequence C: downstream
TAGGAAAGGGGCCAACTAATTGA

>>RXA02225-amino acid sequence
(1-882, translated) 294 residues
QTEERFGAAA DEALAIMLKE ARLQSLLTFV RQLVPAVFSV GLLAYASLLA FDGDITGGEM ISVTLLVPPS
LTVLGVSLGM MTEIWARGQA STKRVQNLVT ELDKAAAEPR PQPATFEFEE GITVWDPSTP EARDVIDREL
EALQVREDVI VAPHRVSVFE GVLKDNLNPM GTIAPEMLRA ALHAASCEDI LSRLGADLNM PGEFELPDTL
IGEAGLNLSG GQRQRIALAR FLAVDPEVLI LDEPTTGLDA VTLDEVAHRV EKLRRGRKTV VITSNPTWHG
VAKQMQSDFS EGVK
>RXA02225-nucleotide sequence B: coding region
CAAACTGAGGAGCGCTTTGGCGCAGCGGCTGATGAAGCCTTGGCAATCATGTTGAAGGAAGCTCGTCTGCAGTCGTT
GCTGACTTTTGTGCGCCAACTTGTCCCAGCGGTGTTTTCTGTGGGTCTTTTGGCTTATGCGTCACTGTTGGCTTTTG
ACGGTGACATAACTGGTGGTGAGATGATCTCGGTGACGTTGCTGGTGCCACCTTCGTTGACTGTGTTGGGTGTGTCG
CTTGGCATGATGACAGAGATTTGGGCTAGGGGACAGGCTTCGACAAAAAGGGTCCAAAACTTAGTCACTGAACTGGA
TAAGGCGGCCGCTGAGCCACGACCTCAGCCTGCCACCTTTGAATTTGAAGAGGGGATCACGGTGTGGGATCCTTCGA
CACCTGAGGCACGCGATGTGATTGATCGGGAGTTGGAGGCGCTTCAGGTTCGCGAAGATGTCATTGTGGCTCCTCAC
CGCGTCAGCGTGTTTGAAGGTGTGCTGAAGGATAATTTGAATCCGATGGGCACTATCGCACCGGAGATGCTGCGCGC
TGCTCTTCATGCTGCAAGTTGTGAGGACATCTTGAGCCGATTGGGTGCTGATCTGAACATGCCGGGGGAGTTTGAGC
TTCCAGATACCTTGATCGGCGAGGCCGGATTGAATCTCTCCGGTGGCCAACGCCAGAGGATTGCTTTGGCACGATTC
TTGGCTGTTGATCCTGAGGTGCTCATTTTGGATGAACCGACCACGGGGTTGGATGCGGTGACCCTGGATGAAGTGGC
ACATCGCGTCGAAAAGCTTCGTCGAGGCCGGAAAACCGTCGTCATTACGTCGAACCCGACGTGGCACGGCGTCGCAA
AGCAGATGCAATCTGATTTTTCGGAAGGGGTGAAG
>RXA02225-nucleotide sequence C: downstream
TAGATGGCGCAGCATGAGCGCGT

>>RXA02749-amino acid sequence
(1-876, translated) 292 residues
MSPILKVRDL VKRYGDTVAV DGLNFDVSQG EIFAFLGENG AGKTTTISCL IGIDQATSGE IELQGGQVDS
EKLGVVFQQS VLDPLLSAKE NLETRGQLYP GVGKQRVAQL IEQIGMEGFA DRRYGVLSGG EKRRTDIARA
LLHSPDILFL DEPTAGLDPR SRRQVWDTIN SLRNDVGLTV FLTTHYMEET ELADSVLIID RGKEVASGTP
MELRARYTTT ELTLRTNDPT HSGKELAHLS PEIDGDRLRI KLENGLEAAR LATELDGVLD VEIRHGSMDD
VFLAVTAERK RS
>RXA02749-nucleotide sequence A: upstream
CAACCTAGACTTCGGTAAGAAGTAACTTTGCTTTAGTTGGTCGGCGCATCACTTTCCCTAAGCGATGCGCCGATTAC
TTGTTTTTGCTACAAATTTAACT
>RXA02749-nucleotide sequence B: coding region
ATGTCACCCATCCTAAAAGTGCGGGACCTCGTCAAACGCTATGGCGACACCGTTGCGGTTGACGGTTTAAATTTTGA
TGTTTCACAAGGGGAAATTTTTGCCTTTCTAGGGGAGAACGGCGCAGGAAAAACAACCACGATTTCATGCCTGATTG
GCATTGATCAAGCCACCTCTGGGGAGATCGAACTGCAGGGTGGCCAAGTAGATTCTGAAAAACTTGGAGTGGTGTTT
CAACAATCCGTCTTAGACCCTTTGCTGAGTGCCAAAGAAAACTTGGAAACACGCGGACAGCTGTACCCAGGGGTGGG
GAAGCAGCGGGTTGCACAGCTCATTGAGCAAATCGGGATGGAAGGGTTTGCGGACCGCCGATACGGAGTGTTGTCGG
GCGGTGAAAAACGTCGCACCGACATCGCACGAGCTTTACTGCACAGCCCAGACATTCTTTTTCTTGATGAACCCACA
GCAGGCCTCGACCCCAGATCACGACGCCAAGTTTGGGACACCATCAATTCCCTGCGTAACGATGTGGGCCTCACTGT
CTTTTTGACCACTCACTACATGGAAGAAACAGAACTGGCTGATTCAGTTCTAATCATTGACCGTGGCAAAGAGGTCG
CATCAGGAACCCCGATGGAACTGCGCGCCCGTTACACCACAACAGAATTGACTCTTAGAACAAACGACCCTACTCAT
TCGGGTAAAGAGTTGGCCCACTTGAGCCCAGAAATCGACGGTGACCGACTGCGGATCAAGTTGGAAAATGGGCTCGA
AGCTGCGCGCCTGGCAACAGAACTAGATGGGGTTCTCGACGTAGAGATCCGCCACGGTTCCATGGACGATGTATTTC
TAGCAGTTACAGCTGAACGGAAACGATCA
>RXA02749-nucleotide sequence C: downstream
TGATTACAGTTCTGACACGCAGA

>>RXA02571-amino acid sequence
(1-1029, translated) 343 residues
VVALTQIVGP SGSGLTRELE KRYRETPGAV MLTADPRAHI TYLRATVAEE LAFGLEQRGI VPAQMWERVR
NIGLGLENLL DRAPAQLSGG QTRRLAIGTV AILEAPTMLL DDPLSGLDTS SRAQLITMLE SYEGDVIVAA
HKRWLDAPTV YLGDLEELSL PARVEFSGPS RTFSAITGTR GQQRRRWWQF NESQPQFQIG PLDITVSAGQ
VLWLQGPNGS GKSTLLRGLA NEPGTELMLQ NPSDQVIDST VANWVPGSNS EEHPLDLSQR ELRLAQCDAA
LGNNPEVLLA DEPDVGLDVG GRNAIHQRFA DFLGNGGALI LTCHDETFVA EVAEYAIVKE MGL
>RXA02571-nucleotide sequence A: upstream
TGGACAGGCCGGGGCCGCGTACGGTGTTGGTTGAGGTGGTGGAGGGGCGCGTCGAAAAGCATTGTCGCTGGTTGTTG
CCGCTTTTGGCAGTCGGGATGGC
>RXA02571-nucleotide sequence B: coding region
GTGGTGGCTCTAACTCAAATCGTCGGACCGTCCGGCTCCGGGCTCACGCGGGAATTGGAAAAACGCTACCGGGAAAC
GCCCGGAGCGGTGATGCTGACCGCCGACCCGCGCGCGCATATCACCTACCTGCGCGCGACAGTCGCCGAGGAGCTGG
CCTTTGGGCTGGAACAACGCGGCATCGTACCCGCGCAGATGTGGGAGCGCGTCCGAAACATCGGGCTCGGCCTCGAG
AATCTGCTAGACCGCGCACCCGCGCAACTTTCCGGCGGGCAAACACGGCGGCTGGCGATCGGCACCGTCGCCATCTT
AGAGGCGCCAACGATGCTTCTCGACGACCCCCTCTCCGGTCTTGATACCTCCTCGCGAGCCCAACTCATCACAATGT
TGGAATCATATGAGGGCGATGTCATCGTCGCTGCGCACAAGCGGTGGCTCGACGCGCCGACTGTGTACTTAGGGGAT
TTGGAGGAGCTGTCCCTGCCTGCGCGGGTGGAATTTTCCGGTCCATCGCGAACGTTTTCAGCGATTACAGGAACCCG
CGGACAACAACGCCGACGCTGGTGGCAATTCAACGAATCCCAACCACAGTTTCAGATCGGCCCCCTGGATATTACTG
TTTCTGCAGGTCAAGTGCTGTGGTTGCAGGGTCCCAATGGTTCAGGGAAGTCCACACTCCTGCGTGGTCTTGCCAAT
GAACCCGGCACTGAATTGATGCTGCAAAACCCTAGCGATCAAGTCATTGACTCCACTGTTGCTAATTGGGTGCCAGG
CAGTAACAGTGAAGAACATCCGCTGGATTTATCGCAACGCGAACTCCGCCTTGCCCAATGCGACGCAGCCCTGGGTA
ATAACCCGGAAGTTTTGCTTGCTGATGAACCCGACGTCGGCCTTGATGTCGGCGGTCGAAACGCCATCCACCAGCGC
TTTGCGGATTTCTTAGGGAATGGGGGAGCGCTGATCCTGACCTGCCATGATGAAACCTTCGTGGCAGAGGTAGCTGA
ATACGCGATAGTGAAGGAAATGGGGCTC
>RXA02571-nucleotide sequence C: downstream
TAGGTTTCTTTGGACCAAACCAC

>>RXA02547-amino acid sequence
(1-2124, translated) 708 residues

AARLTVDEYP AAREALESAG QRNVEDRTRA VDEFKAADQE LSSLSKGSSN IEYRLLQVRE NLCQDLGVSP
RDMPFAGELI DPNNAEWEPV VQRILGGFAA EMLVPHGLLP RVRDWVNAKH LAALLKFNGV VTTGEYKTSR
FPADSLIRKV DVVESPFRDW VNQELGKRFN IRCVRTPEEL SALGPRDQGV TILGVRKFAQ QTGDPTTRWE
KDDRRKLGDR STYRLGSTND AKVETLRETV KAGKAVVQAA DNRIAANRAE LRELERQYQA SQEILKVSWA
QIDVESADAA IAELDRLLEE LNNTPEATEL SARHEAAKQT LARVSDLLVA AQSEETVASM NLKRAETELK
RLESLPVAEV SEEIAREVEK LFLANTRRVH AANVDEQTIA LREDLDKQID ANEAELRRCE NQIVGILRSY
IETWPANRAD LQAEPEFVGE AINRLGELRS DRLAEFTAKF LGLMNEMSTR NLGQISRRLR DARREIEERI
EPINASLAQS EFNEGRFLHI DIRDQSGPIV REFQQKLDAA TSGDLGTSTE KQAFARYALI AEIISKLASH
DSADARWRNT VLDTRRHVRF IGLERDSDGA TVNTYVDSAS LSGGQAQKLV FFCLAAALRY QLAEPGAHYP
TYATVILDEA FDRADPAFTR QTMNVFHSFG FHMVLATPLK LIQTLGDYVG STIVVSYTEK PNAQGAIQGN
SSFSRIEK

>RXA02547-nucleotide sequence B: coding region

GCTGCGCGGCTGACCGTGGATGAGTATCCGGCGGCGAGGGAAGCGCTTGAATCTGCAGGTCAGAGGAATGTAGAGGA
CCGAACCCGTGCGGTTGATGAGTTCAAAGCGGCGGATCAAGAGCTGTCTTCTTTGAGTAAAGGCAGCAGTAATATTG
AGTACCGTTTGCTGCAGGTGCGGGAAAATTTGTGTCAGGATTTGGGCGTGAGCCCGCGGGATATGCCCTTTGCCGGT
GAGCTGATTGATCCGAATAATGCGGAATGGGAACCCGTTGTGCAGCGCATTTTGGGTGGTTTTGCTGCGGAAATGTT
GGTTCCTCATGGGTTGTTGCCACGGGTTCGGGATTGGGTAAATGCCAAACATTTGGCAGCGCTGCTGAAATTCAACG
GCGTGGTGACAACGGGGGAGTACAAAACCTCGCGTTTTCCGGCGGATTCCCTGATCCGAAAAGTTGATGTTGTGGAG
TCGCCGTTTCGCGATTGGGTAAATCAAGAATTAGGCAAGCGTTTTAATATTCGGTGCGTGCGCACTCCTGAGGAATT
GTCGGCGCTGGGGCCACGCGATCAGGGCGTGACCATTTTGGGTGTGCGAAAATTTGCGCAGCAGACAGGCGATCCGA
CGACGCGTTGGGAAAAAGATGATCGCCGAAAGCTGGGGGATCGTTCCACATACCGTTTGGGTTCCACCAATGATGCC
AAGGTGGAAACGCTTCGGGAAACCGTGAAAGCTGGCAAAGCAGTTGTGCAGGCAGCTGATAATCGCATTGCTGCAAA
CCGCGCTGAGCTGCGGGAACTTGAACGGCAGTATCAAGCTTCGCAAGAAATTTTGAAAGTGTCGTGGGCTCAGATTG
ATGTGGAATCAGCCGACGCGGCGATTGCTGAGCTGGACCGATTGCTGGAAGAGCTGAACAACACTCCAGAGGCCACC
GAGCTTTCCGCGCGGCATGAGGCGGCGAAGCAGACGCTCGCGAGGGTTTCTGACTTGCTTGTCGCAGCTCAGAGTGA
GGAAACCGTGGCGTCGATGAACCTGAAACGCGCCGAAACTGAATTGAAACGGCTCGAAAGCCTGCCGGTTGCGGAGG
TTTCTGAAGAAATCGCGCGGGAAGTGGAGAAACTATTTCTTGCCAACACCCGCCGGGTTCACGCCGCCAACGTGGAT
GAGCAGACCATTGCGCTGCGCGAGGATCTGGACAAACAAATCGATGCCAATGAGGCAGAACTTCGACGTTGTGAAAA
CCAAATTGTTGGCATTTTGCGCAGCTATATTGAAACGTGGCCTGCGAACCGCGCTGACTTACAAGCCGAACCTGAGT
TTGTTGGTGAGGCCATCAACCGCCTCGGCGAGCTTCGCAGCGATCGTTTGGCAGAATTCACGGCCAAATTCCTAGGG
CTCATGAACGAGATGTCCACCCGAAACCTCGGCCAAATCTCGCGGCGTCTACGTGATGCGCGCCGGGAAATCGAGGA
GCGCATCGAGCCGATCAACGCCTCCTTGGCGCAGTCGGAATTCAACGAAGGTCGCTTCCTGCACATCGACATCCGTG
ATCAAAGTGGTCCGATTGTGAGGGAATTCCAGCAGAAACTTGATGCCGCTACCAGCGGTGACCTGGGAACCAGTACC
GAGAAACAAGCCTTCGCCCGTTATGCGCTGATCGCTGAAATCATTTCCAAACTCGCCTCCCACGACTCCGCCGACGC
CCGCTGGCGCAACACCGTTCTAGACACCCGCCGCCACGTTCGCTTCATCGGCCTCGAGCGCGATTCCGACGGCGCAA
CCGTCAACACCTACGTCGACTCCGCATCACTTTCAGGCGGACAAGCCCAGAAGCTGGTGTTTTTCTGCCTCGCCGCT
GCCTTGCGCTACCAGCTAGCCGAACCCGGCGCCCATTATCCCACCTACGCCACCGTCATTCTGGACGAAGCCTTCGA
CCGCGCCGACCCCGCCTTCACCCGCCAAACCATGAACGTCTTCCACAGCTTCGGCTTCCACATGGTGCTCGCGACCC
CGCTGAAACTTATCCAAACCCTCGGCGATTATGTCGGCTCCACCATCGTGGTCAGCTACACCGAAAAACCAAACGCC
CAGGGCGCAATTCAGGGCAATTCCAGTTTCTCTAGGATCGAGAAA

>RXA02547-nucleotide sequence C: downstream

TAACATGCCATTGTTTATCGACG

>>RXA01604-amino acid sequence
(1-606, translated) 202 residues
DTPFADVEIA PDSGLTLLST GRESQSSSFS LVLSGRMRAS TGTIELNGEP IKATKLAKHV ALAGIPEIDS
LERLVTVRTV VREQLAWSSP WYLMVPRDIS DSGRWVDVEK HLGLNLNPKT LIGDLSVLER FKLRIALALL
ARPEAQLLVV DDPDQVRSME LRAEVLHALK GVAEDLPVVV VSTNPDFDSL ADTALTITGA GN
>RXA01604-nucleotide sequence B: coding region
GACACACCCTTCGCCGATGTTGAGATAGCTCCAGACAGCGGACTCACTTTGCTGAGCACCGGGCGCGAATCCCAATC
CAGTTCCTTTTCTTTGGTACTTTCCGGCCGCATGCGCGCCTCCACCGGAACCATCGAATTAAACGGCGAACCCATCA
AGGCAACCAAGCTGGCCAAGCATGTGGCTTTGGCGGGCATCCCTGAAATCGATTCACTCGAGCGACTTGTCACTGTG
CGCACCGTTGTCCGTGAACAACTCGCCTGGTCAAGCCCTTGGTACCTGATGGTGCCCAGGGATATTAGTGATTCGGG
ACGGTGGGTTGACGTCGAAAAGCATCTTGGCCTGAACCTGAACCCTAAAACCTTAATCGGCGACCTCAGCGTGCTCG
AGCGTTTTAAGCTGCGCATCGCGCTGGCGCTGCTGGCGCGGCCAGAGGCGCAACTGTTGGTCGTGGATGATCCCGAT
CAAGTGCGCAGCATGGAATTGCGTGCGGAGGTGTTGCACGCATTGAAAGGCGTTGCAGAGGATCTCCCTGTGGTCGT
GGTATCCACCAACCCAGATTTTGATTCCTTGGCCGATACCGCTTTGACCATTACGGGGGCTGGAAAC
>RXA01604-nucleotide sequence C: downstream
TAATGGCATTTTTACACTTTGGC

>>RXA00456-amino acid sequence
(1-312, translated) 104 residues
VLQALLAIMV SLSVAAILEG NRALVGLLLA TTLGLGVAQW IQKVVAEDLG QHYVHEVRRE LVGAALVPGN
TASLGVTVTR ASNDLTAVRN WVALGIVPMV TGLP
>RXA00456-nucleotide sequence A: upstream
CTCACCAACCCGGAGATCGTCACAGCGGTGCTAACGGATCATGCCTAGCTTATGGCGTGCTCGTCGCAGACTTTTGC
TCATTGCCCTAGGTGTACTTGGT
>RXA00456-nucleotide sequence B: coding region
GTGCTGCAGGCACTGCTGGCGATCATGGTGTCGTTGAGCGTAGCCGCCATACTTGAGGGAAACCGAGCACTTGTTGG
ATTGCTGCTTGCTACCACGTTGGGTTTGGGGGTGGCGCAGTGGATTCAAAAAGTAGTGGCAGAAGATCTAGGCCAGC
ATTATGTGCATGAGGTGCGTCGTGAATTGGTGGGTGCTGCGCTGGTGCCTGGAAATACGGCCTCGTTGGGCGTGACT
GTCACCCGAGCCAGCAATGATCTCACCGCGGTGCGCAATTGGGTGGCTTTGGGCATTGTTCCGATGGTCACCGGGCT
GCCG

>>RXA00243-amino acid sequence
(1-927, translated) 309 residues
VTSEQALDPI HPGQFRLSRI QLINWGTFHG TVDIPVTREG ILVTGGSGSG KSTLIDAITA VLLPQGKLRF
NSAAQANTPR NKGRSLVTYI RGAWRAQEDP LQDQIVSTYL RPRATYSLVG LTYSNGEGVE HTLVAIFYLK
SGHNLTSDIS SYYGVFPVDQ DINALLDFLK EGIDKRQIRA AFKEAIFSEQ HSVFSGRFRS RLGISSEEAL
LLLHRAQSAK DLQSLDDLFR DYMLVEPDTF SIAKTAVEQF QDLEGAYEQV EDIKRQIHTL DPLVQLKNRR
EKAQQSKDHA NALKKALPTV GNRIKKEEQ
>RXA00243-nucleotide sequence A: upstream
CACTGCGCCAGATTTTTGATGCCGACACTGTGGCAGGTGTGCGCGCTGAGTACGAAAAATTTAACAAAGCAGCCCAT
GATGGAAATGAAGAGGAACAGAA
>RXA00243-nucleotide sequence B: coding region
GTGACCAGCGAACAAGCTTTAGATCCTATCCACCCAGGTCAGTTCCGTCTTTCTCGGATTCAGTTGATCAACTGGGG
AACCTTCCACGGAACGGTGGACATTCCTGTGACCAGGGAAGGAATCTTAGTTACCGGTGGTTCGGGATCAGGAAAAT
CCACGCTGATTGATGCGATCACGGCGGTATTGCTTCCGCAAGGAAAGCTGAGGTTTAACTCTGCCGCACAGGCTAAT
ACTCCGCGGAATAAGGGACGCAGTTTGGTTACCTATATCCGTGGCGCTTGGCGTGCGCAGGAGGATCCGCTGCAGGA
TCAGATTGTCTCCACGTACCTACGTCCCCGCGCAACCTATTCGCTGGTTGGATTGACTTATTCCAACGGTGAAGGCG
TCGAGCACACCTTGGTGGCTATTTTCTATCTGAAATCGGGACACAATTTAACCTCCGATATTTCTTCATATTATGGT
GTGTTTCCCGTTGATCAAGACATCAATGCGCTGCTGGATTTCCTGAAAGAGGGCATCGATAAACGCCAGATCAGAGC
TGCTTTCAAGGAAGCCATCTTTAGCGAGCAGCATTCTGTATTCTCCGGCAGGTTTAGAAGCCGTTTGGGGATCTCCA
GTGAGGAAGCTTTGCTGTTGTTGCACCGCGCGCAGTCGGCGAAAGATCTTCAAAGCTTGGATGATCTATTTCGGGAT
TACATGCTGGTGGAACCGGATACGTTCAGCATTGCCAAAACTGCCGTGGAACAATTCCAAGACCTTGAAGGTGCTTA
TGAGCAGGTCGAAGATATTAAACGGCAGATCCACACCCTGGATCCTTTGGTGCAGCTGAAGAATCGGCGAGAGAAAG
CGCAACAGTCCAAAGATCATGCCAATGCACTGAAGAAGGCGCTGCCGACTGTCGGGAATCGCATTAAGAAGGAAGAG
CAA

>>RXA00259-amino acid sequence
(1-2202, translated) 734 residues
MSGLFTPFSD AAKNNTVKTD GDSVSGRDLP ITKISEDRFE RSAYSAQLAN IICDVAPWGA STVFSLTGQW
GSGKTSLVNL IRSEESLSNE KWTIVDFNPW VASDPQSLIE EFYRVIVGTV PDDKTGQKIK TVLQKTFSTI
GSIAGGVGGF GVLEALALSK GVDAANAVYK TWKQEQDSWP TLYTRAANHF KDLNKRILIV VDDIDRLHTD
ELALLMKVIR LLGRFPQVNY LLVYEEESLL TTLARSTAVG GSEDDALRFM EKIVQYPFDV PPLTSFQIEK
ELSALFDKLF QGVSLSGDPE DFALVKSRMF DVWEKTLVTP RLLHRFAALL TNWTRIYGSG EVNGVDLTIL
ATIRIVFPSV YKRLSRAKEV LLQGGRTTGS QKPGWEKQLC EGMNNEQMDL LKTMLLFLFP RLSDHPSTRM
HRERGISTEV YFDTYLMFQR PGHVISDEQL DKYLSNADDA MGFVDLINSD DNDMVASVMK KLPLAIDRLD
GEGVRHMAVE VLFTAANGMH DKGRQVRMSG IFSDLYSHAC SILGALPQLP VEQLYEKFFS EMTLNEAAFW
LNQVGERARA CGNDVSGLEL FRKVNIKTEA RILSVLKNQD PSDWDLGPYS LGILAKSSNF SSVLKSLQSG
IEEHQFDVID IGVLFLTTVY SSRQGPSGGA WIDSFQHSLF SRYVPDSLRA ITKSEVDVEL GKIQFTDFSW
EGKRKVVAYA LETGRSDFTR ERLGGYSIAD SIVD

>RXA00259-nucleotide sequence A: upstream
GGCCTTATTAACATACGGCGGTTCTAGCACACAGCGATGGATGATGTGTCCCACCGATATTGCTGCATGTGCCAAGA
TTTAGCACGGTACAGTGCTAGAA

>RXA00259-nucleotide sequence B: coding region
ATGAGCGGACTGTTTACCCCATTTTCAGATGCGGCAAAAAACAACACGGTAAAAACTGATGGAGATTCAGTATCTGG
TCGAGACTTGCCTATTACTAAGATCTCTGAGGATCGTTTCGAGCGTTCTGCGTATTCAGCCCAGCTGGCAAATATAA
TCTGCGATGTGGCACCTTGGGGAGCGAGCACTGTTTTCAGTCTTACTGGTCAGTGGGGCAGTGGTAAGACATCTCTT
GTTAATTTGATTCGCTCGGAAGAATCTCTATCGAACGAAAAATGGACAATCGTTGATTTCAACCCGTGGGTGGCCTC
TGACCCGCAATCTTTGATTGAGGAGTTTTACCGAGTAATCGTTGGGACGGTACCTGATGATAAGACCGGCCAAAAGA
TCAAAACTGTTCTGCAGAAAACCTTTAGCACGATTGGGTCAATTGCAGGTGGGGTCGGAGGGTTTGGTGTCCTAGAA
GCACTTGCGCTCTCAAAAGGAGTAGATGCTGCAAACGCTGTATATAAGACATGGAAACAGGAGCAAGATTCGTGGCC
AACGCTGTATACACGTGCTGCGAACCATTTTAAAGATCTGAACAAGCGAATTCTCATTGTCGTCGATGATATTGATC
GCCTCCATACTGATGAATTGGCGCTGTTAATGAAAGTAATACGCTTGCTTGGACGATTCCCGCAGGTGAATTATCTT
TTGGTTTATGAAGAAGAATCACTGTTAACGACGCTAGCCAGATCGACAGCTGTAGGTGGTAGCGAAGATGATGCTTT
GCGTTTCATGGAGAAAATCGTGCAGTATCCTTTCGATGTTCCGCCTCTGACATCATTTCAAATAGAGAAAGAGCTCA
GTGCATTATTTGACAAGCTTTTCCAGGGTGTTTCGCTATCGGGTGATCCTGAAGACTTTGCACTAGTGAAGTCGAGA
ATGTTCGATGTCTGGGAAAAGACTCTGGTCACGCCGAGGCTGTTGCACCGTTTTGCTGCTCTACTAACCAACTGGAC
TCGGATATATGGATCAGGTGAAGTTAACGGCGTTGATCTCACAATACTTGCGACCATTCGAATTGTTTTTCCGTCTG
TGTATAAACGTCTTTCTCGAGCGAAGGAAGTATTGCTTCAAGGAGGTCGAACGACAGGCTCGCAGAAACCCGGTTGG
GAAAAGCAATTATGTGAGGGGATGAACAACGAGCAGATGGATCTTTTAAAGACCATGCTTTTGTTCCTTTTCCCACG
TCTTTCGGATCACCCTAGTACGAGAATGCATCGTGAGAGGGGGATCTCGACGGAAGTTTATTTTGACACGTACCTCA
TGTTTCAAAGACCTGGACATGTCATAAGTGATGAACAGTTGGATAAGTATCTATCTAATGCGGACGATGCTATGGGT
TTCGTCGATTTAATTAACTCCGATGACAATGACATGGTGGCATCAGTGATGAAAAAGCTTCCTCTAGCAATTGATCG
ACTTGATGGAGAGGGTGTTAGGCACATGGCAGTTGAGGTGTTATTCACCGCTGCTAATGGTATGCATGATAAAGGTC
GTCAAGTGCGTATGAGCGGCATATTCAGTGACCTGTATTCCCATGCGTGCTCGATTCTTGGTGCATTGCCTCAATTA
CCAGTGGAACAACTCTATGAGAAATTCTTTTCTGAGATGACGCTTAATGAGGCTGCTTTCTGGTTAAACCAGGTGGG
GGAAAGGGCTAGAGCCTGTGGTAATGATGTAAGTGGCCTTGAGCTTTTTCGTAAAGTTAATATAAAGACCGAAGCTA
GAATTTTAAGTGTATTGAAGAATCAGGACCCCTCAGATTGGGATTTAGGTCCATATTCGCTTGGTATTTTGGCGAAA
AGCTCGAATTTTTCTTCAGTGCTGAAGTCTCTGCAAAGTGGTATAGAGGAACATCAGTTTGATGTGATAGATATTGG
AGTGCTTTTCTTAACGACTGTGTATTCTTCGCGACAGGGACCAAGCGGTGGTGCATGGATAGATTCTTTTCAGCATA
GTCTGTTTTCACGGTACGTACCTGATTCTCTACGGGCTATAACCAAGTCTGAAGTAGATGTAGAACTAGGTAAGATA
CAGTTCACGGATTTTAGCTGGGAAGGGAAGCGAAAAGTTGTCGCATATGCACTGGAGACTGGAAGAAGTGATTTCAC
TCGAGAACGATTAGGGGGCTACAGTATCGCAGATTCTATAGTCGAT

>RXA00259-nucleotide sequence C: downstream
TGATGAGGCTGAGGTCATGACTT

>>RXA00164-amino acid sequence
(1-1689, translated) 563 residues
VGRIPRAKWW FLGALVLLSA GAYASVLVPQ VLGRIVDLVS DGAQMRDFVE LSVILIAVAI AGAVLSACGF
YVVSRISEKI IANLREDMVG TALGLPTHQV EDAGSGDLVS RSTDDVSELS AAVTETVPIL SSSLFTIAAT
IIALFSLDWQ FVLIPVVVAP VYYFASKHYL SKAPDRYAAE RAAMAERARK VLEAIRGRAT VRAYSMEDAM
HNQIDQASWS VVVKGIRART TMLILNMWML FAEFLMLAVA LVIGYKLVID NALTIGAVTG AVLMIIRLRG
PMNMFMRVLD TIQSGYASLA RIVGVVADPP IPVPDSGVKA PQGKVELRNV SFSYGDSWAV KDIDITINSG
ETVALVGASG AGKTTVAALL AGLRVPDQGQ VLVDDFPVSH LSDRERIARL AMVSQEVHVF SGTLRQDLTL
AKPDASDEEL AHALGQVNAL DWLESLPEGL DTVVGARGIQ LEPVVAQQLA LARVLLLNPA IVIMDEATAE
AGSAGASALE EAADAVSKNR SALVVAHRLD QASRADQILV MDKGEVVESG THQELLDHGG IYQRLWTAWS
VGR
>RXA00164-nucleotide sequence A: upstream
CTGCTTTGCGGGAGGTTATGAAATGAGTGGGGAGACGTCGAAAAGCATGCGCTTTCCGTTGGCCAGCCTGCCGCAAG
TGCGGCGCGAGGTGGCCCGGCAG
>RXA00164-nucleotide sequence B: coding region
GTGGGTCGTATTCCGCGGGCGAAGTGGTGGTTTTTAGGCGCGCTGGTGTTGCTGAGTGCGGGCGCTTATGCGTCGGT
GCTGGTGCCGCAGGTGCTGGGGCGGATTGTGGATCTGGTGTCCGATGGCGCGCAGATGCGTGATTTTGTTGAGCTCA
GTGTGATTCTCATTGCGGTGGCAATTGCCGGCGCGGTGCTCAGTGCGTGCGGGTTCTATGTGGTGTCGCGGATTTCT
GAGAAGATTATCGCCAATTTGAGGGAAGATATGGTGGGCACCGCGCTTGGGTTGCCCACGCACCAGGTGGAAGATGC
GGGCTCTGGCGATTTGGTGAGCCGCTCCACCGATGATGTCTCCGAGCTATCCGCAGCGGTGACAGAGACCGTCCCGA
TTTTAAGTTCCTCACTGTTTACCATTGCCGCGACGATCATTGCGCTGTTTTCTTTGGACTGGCAATTTGTGCTCATT
CCTGTCGTGGTGGCGCCGGTGTACTACTTCGCGTCCAAGCACTATTTGAGCAAGGCGCCGGATCGGTATGCGGCAGA
ACGCGCGGCGATGGCGGAGCGTGCGCGAAAGGTACTTGAGGCTATTCGCGGGCGTGCAACTGTGCGGGCGTATTCCA
TGGAAGATGCCATGCATAATCAGATTGATCAGGCGTCGTGGTCTGTGGTGGTCAAGGGTATTCGTGCGCGCACCACC
ATGTTGATTTTGAACATGTGGATGCTGTTTGCGGAATTCCTCATGCTCGCGGTCGCGTTGGTGATCGGCTACAAGCT
GGTCATTGATAATGCGCTGACGATCGGCGCGGTTACCGGTGCCGTGCTGATGATTATTCGTCTGCGTGGCCCGATGA
ATATGTTCATGCGCGTGCTCGACACCATTCAATCCGGCTATGCGTCGCTGGCGCGCATCGTGGGAGTTGTTGCGGAT
CCGCCGATTCCTGTGCCCGACAGCGGTGTGAAAGCACCTCAGGGCAAAGTGGAATTGCGCAACGTCAGCTTTAGCTA
TGGCGATTCCTGGGCGGTGAAAGACATCGACATCACGATCAATTCCGGCGAAACTGTCGCGCTCGTGGGCGCATCTG
GCGCAGGTAAGACGACGGTCGCCGCCTTGCTGGCGGGCTTGCGGGTGCCAGATCAAGGGCAAGTGCTTGTCGACGAC
TTCCCCGTCTCTCACCTCTCTGACCGCGAGCGTATCGCCCGCTTGGCCATGGTCAGCCAGGAGGTTCATGTTTTCTC
CGGCACGCTGCGCCAGGATCTCACCTTGGCTAAACCAGATGCCTCCGATGAGGAATTAGCGCATGCTCTTGGGCAAG
TTAATGCCCTTGACTGGTTGGAGAGTCTTCCAGAAGGACTGGACACGGTCGTTGGTGCGCGAGGAATCCAGCTAGAA
CCAGTGGTGGCTCAGCAGTTGGCGTTGGCCCGGGTGTTGTTGCTCAATCCGGCGATCGTCATCATGGATGAAGCCAC
GGCAGAAGCAGGATCGGCGGGTGCCAGCGCACTGGAAGAGGCTGCAGATGCAGTGAGCAAGAACCGTTCCGCATTGG
TGGTGGCGCACCGGTTGGATCAGGCATCGCGGGCTGATCAGATTCTGGTGATGGATAAGGGGGAGGTTGTGGAATCC
GGTACTCACCAGGAGTTATTGGATCACGGGGGTATTTATCAGCGTCTGTGGACTGCGTGGAGTGTCGGAAGA
>RXA00164-nucleotide sequence C: downstream
TAGTTGACTGTTCAATGCGTTGA

>>RXA00165-amino acid sequence
(1-1416, translated) 472 residues
VASAGMAASF ICNGLTPVIV GKAVDEAIGT SDLQRLWFWI AMLAVLFLTA MTVNWIARYM LVRSQQLVSH
DLRMLVTDRI QDPRGFAGKE RTAGGLLSIA SSDTQRVGDI VMMTVFPVAE LASIIYGAVV MYSINPWLSV
AVLIGGPLLV VVAIQVSKPL QKRSGARQQA VAQAAATATD VVQGLRILKG LGAIVTVRRR YEAISGEAYR
KTVHADAAEA RLNGVTDAAG AIFVSALGIG AGFLALQGQM SIGDLITVVG LTQFLIMPMT MLGRNVASRW
ASAEASAKRI RGVLGADFER VSAHDADKAE EIIQQLAKGL TVIRGTDEQL VEVLEQLPRT RVIVAPHAAD
LFDQSVRDNV HPVAEVAEKA IEVASCDDIP GGSSKIVGEG GRLLSGGQRQ RVALARAIAF DPEVLVLQDP
TTAVDSVTEQ NIAQQVAAHR AGKVTIVFSE APAWSAVADQ HVEAAALREV MK
>RXA00165-nucleotide sequence A: upstream
AAACCTCCCCGGGCCCGGCGCGCGACCGTCCAAGATGCCGGCGTTGGATGCCAAATTATGGACTCTCAAAGTGGCGT
TGTCGCAGCGGCCGTGGAGCTTT
>RXA00165-nucleotide sequence B: coding region
GTGGCGTCTGCTGGCATGGCGGCGTCTTTTATCTGCAATGGGTTAACGCCTGTGATTGTGGGTAAGGCGGTGGATGA
GGCTATTGGCACGAGCGATCTGCAGCGATTGTGGTTCTGGATTGCCATGTTGGCGGTTCTTTTCTTAACGGCGATGA
CGGTGAACTGGATTGCTCGGTACATGTTGGTGCGGAGCCAGCAGTTGGTCAGCCATGATTTGCGCATGTTGGTGACT
GATCGGATTCAAGATCCGCGTGGTTTTGCTGGAAAAGAGCGCACTGCGGGTGGATTGTTGTCGATTGCGTCATCGGA
TACGCAGCGGGTGGGCGATATCGTCATGATGACGGTGTTCCCGGTGGCGGAATTGGCGTCGATTATTTATGGCGCCG
TGGTGATGTACAGCATTAATCCGTGGTTGAGTGTGGCTGTGCTGATTGGTGGACCGCTGCTGGTTGTGGTGGCTATT
CAGGTCTCAAAGCCGTTGCAGAAGCGTTCGGGTGCTCGTCAGCAGGCGGTGGCACAGGCTGCGGCTACTGCAACTGA
TGTGGTGCAGGGCTTGAGAATTTTGAAGGGTTTGGGCGCGATTGTCACGGTGCGCCGTCGGTACGAGGCGATTTCTG
GTGAGGCTTATCGGAAGACGGTTCATGCGGATGCTGCGGAAGCTCGCTTGAATGGTGTCACCGATGCGGCGGGCGCC
ATCTTTGTGTCCGCGTTGGGTATTGGAGCAGGATTTTTGGCGCTGCAAGGTCAGATGAGTATTGGTGATTTGATCAC
GGTTGTGGGACTCACACAGTTTTTGATCATGCCGATGACCATGCTTGGTCGAAATGTGGCATCGCGCTGGGCATCGG
CGGAGGCGTCGGCAAAGCGTATTAGGGGAGTGCTCGGTGCTGATTTTGAGAGAGTGTCTGCGCATGATGCGGACAAG
GCTGAGGAGATTATCCAACAACTTGCCAAAGGTTTGACGGTTATTCGAGGCACTGATGAGCAGCTCGTTGAGGTATT
AGAGCAGTTGCCACGTACTCGGGTGATTGTGGCTCCTCATGCGGCGGATCTTTTTGATCAAAGTGTCAGGGACAATG
TGCATCCCGTGGCAGAGGTCGCGGAGAAAGCCATTGAAGTTGCCTCCTGTGACGATATTCCAGGTGGTAGTTCCAAG
ATTGTGGGCGAGGGTGGACGGTTGCTCTCGGGTGGTCAGCGTCAGCGCGTTGCACTGGCTCGGGCGATTGCTTTTGA
TCCAGAGGTGTTGGTGCTTCAAGATCCCACAACGGCAGTGGATTCTGTGACGGAGCAAAACATTGCTCAGCAAGTGG
CAGCACACCGTGCAGGAAAAGTGACCATTGTGTTTAGTGAGGCACCCGCGTGGAGTGCGGTGGCTGATCAACACGTT
GAGGCAGCTGCTTTGCGGGAGGTTATGAAA
>RXA00165-nucleotide sequence C: downstream
TGAGTGGGGAGACGTCGAAAAGC

>>RXA00188-amino acid sequence
(1-1443, translated) 481 residues

LGEFIYTMKN VRKAIGDKVI LDNVYMSFYP GAKIGVVGPN GAGKSSILKI MAGLDQPSNG EAFLDPGATV
GILLQEPPLN EEKTVRQNVE EGLGEIFEKK QRFDAIAEEM ATNYTDELME EMGVLQEALD AADAWEIDSK
IDQAMDALRC PPSDEPVTHL SGGERRRVAL AKLLLAEPDL LLLDEPTNHL DAESVLWLEK HLADYKGAVL
AVTHDRYFLD HVAQWICEVD RGQLHPYEGN YSTYLEKKAE RLEVSGKKDQ KLQKRLKDEL AWVRSGAKAR
QAKNKARLQR YDEMVAEAEK YRKLDFEEIQ IPTPPRLGNK VVEVKDLEKG FDGRVLIKDL SFTLPRNGIV
GVIGPNGVGK STLFKTIVGL ENPDAGSVEI GDTVQLSYVD QGRENIDPEK TVWETVSDGL DYIIVGQNEM
PSRAYLSAFG FKGADQQKPS KVLSGGERNR LNLALTLKQG GNLILLDEPT NDLDVETLGS L

>RXA00188-nucleotide sequence A: upstream

TGCTGTAAGGCCCACTTGCCCGAGGTCCCTAAATTGGCGCTGCTCATGTAATCTTGTGCGTTGAGTACTTAAACGCA
AATTCGCATTTAAAGGGGTTACA

>RXA00188-nucleotide sequence B: coding region

TTGGGCGAGTTCATCTACACGATGAAGAACGTGCGTAAGGCAATCGGTGACAAGGTCATCTTGGACAATGTCTACAT
GTCCTTCTACCCAGGCGCCAAGATCGGTGTCGTCGGACCCAACGGTGCCGGTAAGTCATCGATCTTGAAGATCATGG
CTGGGTTGGATCAGCCTTCCAACGGTGAAGCATTCCTGGATCCAGGTGCAACCGTGGGCATCCTGTTGCAGGAGCCA
CCACTAAACGAAGAAAAGACTGTTCGCCAGAACGTCGAAGAAGGTCTTGGTGAGATCTTCGAAAAGAAGCAGCGCTT
CGACGCCATCGCAGAAGAAATGGCAACCAACTACACCGACGAGCTCATGGAAGAAATGGGCGTTCTGCAGGAGGCTC
TCGACGCAGCTGATGCGTGGGAGATCGACTCCAAGATCGACCAGGCAATGGATGCATTGCGTTGCCCTCCAAGCGAT
GAGCCAGTTACCCACCTCTCCGGTGGTGAGCGTCGCCGCGTCGCACTGGCAAAGCTTCTGCTTGCAGAGCCAGATCT
GCTGCTGCTCGATGAGCCTACTAACCACCTTGACGCCGAGTCCGTCCTGTGGCTGGAGAAGCACCTCGCTGACTACA
AGGGCGCTGTCCTTGCCGTCACACACGACCGTTACTTCCTCGACCACGTTGCACAGTGGATCTGTGAAGTTGACCGT
GGTCAGCTGCACCCTTACGAAGGCAACTACTCCACCTACCTGGAGAAGAAGGCAGAGCGCCTCGAAGTTTCTGGCAA
GAAGGATCAGAAGCTGCAGAAGCGCCTGAAGGACGAACTCGCATGGGTTCGTTCCGGCGCTAAGGCACGTCAGGCTA
AGAACAAGGCTCGTCTGCAGCGTTACGACGAGATGGTTGCGGAAGCTGAGAAGTACCGCAAGCTCGACTTCGAAGAA
ATTCAGATCCCAACCCCACCACGCCTGGGCAACAAGGTCGTGGAGGTCAAGGACCTGGAGAAGGGCTTCGATGGCCG
CGTGCTGATCAAGGACCTGTCCTTCACCTTGCCTCGTAACGGCATCGTCGGCGTGATCGGACCTAACGGTGTGGGTA
AGTCCACCCTGTTCAAGACCATCGTTGGCCTTGAAAACCCAGACGCAGGCAGCGTTGAAATCGGCGACACCGTCCAG
TTGTCCTACGTGGACCAGGGCCGTGAAAACATTGATCCAGAAAAGACCGTTTGGGAAACCGTTTCCGACGGACTCGA
CTACATCATTGTCGGACAGAACGAAATGCCATCCCGCGCATACCTGTCCGCCTTCGGATTCAAGGGCGCAGATCAGC
AGAAGCCATCCAAGGTCCTCTCCGGTGGTGAGCGCAACCGCCTCAACCTCGCGCTGACCTTGAAGCAGGGCGGAAAC
CTGATCCTCCTCGATGAGCCTACAAACGACCTCGACGTGGAAACTCTGGGCTCCCTG

Claims

1. An isolated nucleic acid molecule from *Corynebacterium glutamicum* encoding an MCT protein, or a portion thereof.

2. The isolated nucleic acid molecule of claim 1, wherein said nucleic acid molecule encodes an MCT protein involved in the production of a fine chemical.

3. An isolated *Corynebacterium glutamicum* nucleic acid molecule selected from the group consisting of those sequences set forth in Appendix A, or a portion thereof.

4. An isolated nucleic acid molecule which encodes a polypeptide sequence selected from the group consisting of those sequences set forth in Appendix B.

5. An isolated nucleic acid molecule which encodes a naturally occurring allelic variant of a polypeptide selected from the group of amino acid sequences consisting of those sequences set forth in Appendix B.

6. An isolated nucleic acid molecule comprising a nucleotide sequence which is at least 50% homologous to a nucleotide sequence selected from the group consisting of those sequences set forth in Appendix A, or a portion thereof.

7. An isolated nucleic acid molecule comprising a fragment of at least 15 nucleotides of a nucleic acid comprising a nucleotide sequence selected from the group consisting of those sequences set forth in Appendix A.

8. An isolated nucleic acid molecule which hybridizes to the nucleic acid molecule of any one of claims 1-7 under stringent conditions.

9. An isolated nucleic acid molecule comprising the nucleic acid molecule of any one of claims 1-8 or a portion thereof and a nucleotide sequence encoding a heterologous polypeptide.

10. A vector comprising the nucleic acid molecule of any one of claims 1-9.

11. The vector of claim 10, which is an expression vector.

12. A host cell transfected with the expression vector of claim 11.

13. The host cell of claim 12, wherein said cell is a microorganism.

14. The host cell of claim 13, wherein said cell belongs to the genus *Corynebacterium* or *Brevibacterium*

15. The host cell of claim 12, wherein the expression of said nucleic acid molecule results in the modulation in production of a fine chemical from said cell.

16. The host cell of claim 15, wherein said fine chemical is selected from the group consisting of: organic acids, proteinogenic and nonproteinogenic amino acids, purine and pyrimidine bases, nucleosides, nucleotides, lipids, saturated and unsaturated fatty acids, diols, carbohydrates, aromatic compounds, vitamins, cofactors, and enzymes.

17. A method of producing a polypeptide comprising culturing the host cell of claim 12 in an appropriate culture medium to, thereby, produce the polypeptide.

18. An isolated MCT polypeptide from *Corynebacterium glutamicum*, or a portion thereof.

19. The polypeptide of claim 18, wherein said polypeptide is involved in the production of a fine chemical production.

20. An isolated polypeptide comprising an amino acid sequence selected from the group consisting of those sequences set forth in Appendix B.

21. An isolated polypeptide comprising a naturally occurring allelic variant of a polypeptide comprising an amino acid sequence selected from the group consisting of those sequences set forth in Appendix B, or a portion thereof.

22. The isolated polypeptide of any of claims 18-21, further comprising heterologous amino acid sequences.

23. An isolated polypeptide which is encoded by a nucleic acid molecule comprising a nucleotide sequence which is at least 50% homologous to a nucleic acid selected from the group consisting of those sequences set forth in Appendix A.

24. An isolated polypeptide comprising an amino acid sequence which is at least 50% homologous to an amino acid sequence selected from the group consisting of those sequences set forth in Appendix B.

25. A method for producing a fine chemical, comprising culturing a cell containing a vector of claim 12 such that the fine chemical is produced.

26. The method of claim 25, wherein said method further comprises the step of recovering the fine chemical from said culture.

27. The method of claim 25, wherein said method further comprises the step of transfecting said cell with the vector of claim 11 to result in a cell containing said vector.

28. The method of claim 25, wherein said cell belongs to the genus *Corynebacterium* or *Brevibacterium*.

29. The method of claim 25, wherein said cell is selected from the group consisting of: *Corynebacterium glutamicum, Corynebacterium herculis, Corynebacterium lilium, Corynebacterium acetoacidophilum, Corynebacterium acetoglutamicum,*

Corynebacterium acetophilum, Corynebacterium ammoniagenes, Corynebacterium fujiokense, Corynebacterium nitrilophilus, Brevibacterium ammoniagenes, Brevibacterium butanicum, Brevibacterium divaricatum, Brevibacterium flavum, Brevibacterium healii, Brevibacterium ketoglutamicum, Brevibacterium ketosoreductum, Brevibacterium lactofermentum, Brevibacterium linens, Brevibacterium paraffinolyticum, and those strains set forth in Table 3.

30. The method of claim 25, wherein expression of the nucleic acid molecule from said vector results in modulation of production of said fine chemical.

31. The method of claim 25, wherein said fine chemical is selected from the group consisting of: organic acids, proteinogenic and nonproteinogenic amino acids, purine and pyrimidine bases, nucleosides, nucleotides, lipids, saturated and unsaturated fatty acids, diols, carbohydrates, aromatic compounds, vitamins, cofactors, and enzymes.

32. The method of claim 25, wherein said fine chemical is an amino acid.

33. The method of claim 32, wherein said amino acid is drawn from the group consisting of: lysine, glutamate, glutamine, alanine, aspartate, glycine, serine, threonine, methionine, cysteine, valine, leucine, isoleucine, arginine, proline, histidine, tyrosine, phenylalanine, and tryptophan.

34. A method for producing a fine chemical, comprising culturing a cell whose genomic DNA has been altered by the inclusion of a nucleic acid molecule of any one of claims 1-9.

CORYNEBACTERIUM GLUTAMICUM GENES ENCODING PROTEINS INVOLVED IN MEMBRANE SYNTHESIS AND MEMBRANE TRANSPORT

Abstract of the Disclosure

Isolated nucleic acid molecules, designated MCT nucleic acid molecules, which encode novel MCT proteins from *Corynebacterium glutamicum* are described. The invention also provides antisense nucleic acid molecules, recombinant expression vectors containing MCT nucleic acid molecules, and host cells into which the expression vectors have been introduced. The invention still further provides isolated MCT proteins, mutated MCT proteins, fusion proteins, antigenic peptides and methods for the improvement of production of a desired compound from *C. glutamicum* based on genetic engineering of MCT genes in this organism.